J Grid Computing (2019) 17:279–312
https://doi.org/10.1007/s10723-019-09482-2

What's Happening Around the World? A Survey and Framework on Event Detection Techniques on Twitter

Zafar Saeed · Rabeeh Ayaz Abbasi · Onaiza Maqbool · Abida Sadaf · Imran Razzak · Ali Daud · Naif Radi Aljohani · Guandong Xu

Received: 13 May 2018 / Accepted: 18 March 2019 / Published online: 28 May 2019

Abstract In the last few years, Twitter has become a popular platform for sharing opinions, experiences, news, and views in real-time. Twitter presents an interesting opportunity for detecting events happening around the world. The content (tweets) published on Twitter are short and pose diverse challenges for detecting and interpreting event-related information. This article provides insight into ongoing research. It explores recent research trends and techniques for event detection using Twitter data. We classify techniques and methodologies according to event types, orientation of content, event detection tasks, their evaluation, and common practices. We highlight the limitations of existing techniques and accordingly propose solutions to address the shortcomings. We propose a framework called EDoT based on the research trends, common practices, and techniques used for detecting events on Twitter. EDoT can serve as a guideline for developing event detection methods, especially for researchers who are new in this area. We also describe and compare data collection techniques, the effectiveness and shortcomings of various Twitter and non-Twitter-based features, and discuss various evaluation measures and benchmarking methodologies. Finally, we discuss the trends, limitations, and future directions for detecting events on Twitter.

Keywords Events · Event detection · Twitter · Social media · Machine learning · Survey · Taxonomy · Framework

Zafar Saeed · Rabeeh Ayaz Abbasi (✉) · Onaiza Maqbool
Department of Computer Science, Quaid-i-Azam University, Islamabad, Pakistan
e-mail: rabbasi@qau.edu.pk

Zafar Saeed · Imran Razzak · Guandong Xu
University of Technology Sydney, Sydney, Australia

Ali Daud
International Islamic University, Islamabad, Pakistan

Naif R. Aljohani
Department of Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia

Abida Sadaf
Institute of Information Technology, Quaid-i-Azam University, Islamabad, Pakistan

1 Introduction

Event detection highlights significant happenings in real life. In the late 90s event detection was addressed under the umbrella of Topic Detection and Tracking (TDT). Traditional media content (news stories) was used to analyze and highlight prominent events [7, 114]. The availability of online content has further motivated researchers to explore and analyze these data. Event detection using such content has become a hot research trend because the data that are published online reflect the opinions and experiences of people participating in real-life events [43, 66, 112]. It opens significant opportunities for research in natural language processing (NLP) and machine learning [19, 86]. Event detection systems contribute to reduce the efforts needed for information-seeking tasks regarding world-wide breaking events and provide insight into the opinions of people.

A large amount of user-generated content in different forms is being shared and used online through social media platforms [107, 117]. Social media has exploded as a form of online communication where people from all walks of life can express their thoughts and opinions, communicate with each other anytime and anywhere, creating contents of different types (including bookmarks, photos, and videos) and share them [12, 101]. Social media has a great influence on the social development of a large audience around the world [85, 101]. For instance, social media including Twitter played a vital role in the Arab Spring, with approximately two million tweets a day being posted during the protests [5]. Because of the large user population, these digital online resources become more interesting and motivate researchers to explore what is happening around the world? Information sharing and seeking are common on social media for certain events occurring in the real world. According to Troncy et al., "events are a natural way for referring to any observable occurrence grouping persons, places, times, and activities that can be described" [102]. So, the information related to events is often documented by people. Many users share an enormous amount of information through social media and networking channels in the form of text, images, and videos. Event-based information can appear in several forms such as news, documentary videos, status updates, and images taken before, during and after certain events. The upcoming events shared on such social media also reveal structured information such as title, description, time, and location which is important for analyzing and aggregating event-based information [16].

Of the social media services such as MySpace, Facebook, Bebo, Flickr, del.icio.us, and YouTube, Twitter is well-known because people share their concerns and views regarding events in real-time, generating thousands of tweets per second [82]. Twitter's content contains social information and temporal characteristics, and analyzing such content can reveal valuable information [13]. Twitter acts as a real-time, diverse and dynamic content publisher and is well-studied for event analysis. The identification of escalated events can also be useful for government and non-government organizations, for example, detecting and identifying natural events, such as earthquakes, flood and rain in time can help to accelerate support activities. It can also effectively contribute to help state institutions in efficient decision-making and policy-making after analyzing recent events of interest [36, 50], such as traffic jams, security threats, and epidemics within a specific geographical region.

Twitter was launched in 2006 [56]. The core feature of Twitter is to allow users to post and read short messages called tweets to share information, thoughts, opinions and ideas which inform people about what is happening right now? Twitter is the most famous and fastest growing micro-blogging service [13, 19]. According to Alexa traffic rankings,[1] Twitter is ranked in 13th position for global usage. Other official company statistics[2] show that it has 313 million monthly active users, however the total number of registered users is not disclosed. 500 million Tweets are sent per day (i.e. $\approx$ 5800 tweets per seconds). 82% of active Twitter users are on mobile phones and 79% of the total accounts are outside the U.S. Due to the large number of users from different countries, Twitter supports more than 40 different languages to create and publish short content.

A user on Twitter can post tweets which become instantly accessible to their followers and public (if allowed by the user). The majority of users tweet through mobile apps, resulting in the instant dissemination of information. Various features have added in Twitter, these being real-time, easy-to-use, and portable. This micro-blogging service has a unique set of features that make it more interesting than other social networking and blogging services. Unlike Facebook, the users' network is a directed graph (i.e. asymmetric relation) on Twitter. A registered user can follow any other user and is called "Follower". The user who is being followed is called the "Followee".

[1] http://www.alexa.com/siteinfo/twitter.com (accessed on January 28, 2019)

[2] http://files.shareholder.com/downloads/AMDA-2F526X/5439610324x0x961126/1C3B5760-08BC-4637-ABA1-A9423C80F1F4/Q317_Selected_Company_Metrics_and_Financials.pdf (accessed on January 28, 2019)

Followers receive tweets whenever a new message is posted by a followee, but not vice versa.

Users can share someone's tweet using a ***retweet***. Whenever a tweet is re-tweeted, Twitter formulates the underlying content by adding "RT" followed by "@UserName" which is the person who generated the actual tweet [21]. There are two other ways to retweet, i.e., auto-retweet and with comment. In an auto-retweet, the contents of the original tweet are preserved and posted along with the retweeting user's name. Users can also add their comments beside the original tweet. A ***mention*** is a way to reply to a registered user. To mention another user, the "@" sign followed by the "Screen Name" is used anywhere inside the tweet content. A ***hashtag*** is a well-known concept in social media, also known as an explicit content descriptor. It is a word or phrase without spaces, starts with a "#" symbol often used within the tweet to highlight topic(s) [58]. External URLs can also be added to a tweet when detailed information about a topic is required. Twitter allows its users to ***like*** tweets. Selecting the *like* option in a tweet signals the original user that someone liked their tweet. Other silent features include *geo-tag* (if enabled by the publisher) and *tweet time*, among many others [39].

Using these features or a subset, recent studies have created novel research applications to detect events in various domains, such as natural disaster emergencies [33, 93, 100, 108], emerging political events [78, 80, 88, 91, 92], sports [30, 62, 92], traffic events and conditions [87], epidemic diseases [59], and show business [30].

Given the importance of Twitter as a social media platform, and its role in enabling the detection of events, in this paper, we present a survey of event detection research based on Twitter. Most of the existing research surveys are for event detection using social multimedia content [69, 104]. Twitter is one of the most popular micro-blogging platforms, as it has become a medium for the real-time social broadcasting of ongoing events. However, we did not find an extensive research survey that addresses event detection research based on Twitter. We fill this vacuum by creating an event detection taxonomy which is orthogonal to the hierarchy proposed by Atefeh et al. [13]. Our survey provides a comprehensive look into event detection methods and their application domains. In addition to the taxonomy, our work covers broader perspectives such as data collection techniques, evaluation techniques and measures, event detection tools, comparative and critical discussion on research trends and common practices, and feature design. Specifically, our survey covers the following aspects:

- Critical analyses and discussions on event detection techniques
- Comparison of dataset collection strategies, benchmarking and evaluation measures.
- Existing tools and systems for event detection.
- Challenges and proposed solutions.
- Discussions about research trends, limitations, and future directions.
- A generic *Event Detection on Twitter* (EDoT) framework.

The rest of the paper is organized as follows: Section 2 lists the concepts and terminology required for understanding the topic of event detection. The data collection strategies which are necessary for all event detection methods are discussed in Section 3. It also provides a comparative analysis of Twitter APIs used for crawling the data, a comparative analysis of temporal coverage and size, benchmarking, and evaluation. Section 4 discusses features related to tweets. Event detection techniques and the related work are discussed in Section 5. The available tools and systems are discussed in Section 6. A critical analysis of research trends, challenges and future directions is discussed in Section 7. Finally, Section 8 concludes the paper.

2 Concepts and Terminologies

In this section, we provide descriptions of concepts and terminologies related to the area of event detection.

2.1 Event: Definition and Context

The concept of an event varies across several disciplines. There is a lack of a formal and standardized definition of an event. In the following paragraphs, we start with a philosophical definition of an event. Next, we provide definitions of an event from literature in the context of social media.

According to the Stanford Encyclopedia of Philosophy,[3] events are - “Things that happen”, like births, deaths, thunders, lightning, and weddings. In the real world with the human perspective, “the intention to plan and execute actions, and to bring about changes in the world” is considered as an event.

According to the studies in late 90s [6, 7, 114], an event is defined as “something that happens at a specific time and place with consequences”. The outcome and consequences motivate people to perform certain actions on social media hence disseminate event related information in online social networks. A similar definition is given by McMinn et al. [75] that an event is “something significant that happens at specific time and place”. Social media users post online information about significant happenings. In this context, Weng and Lee [112] describe an event as “a set of posts sharing the same topic and words within a short time”. The definition is formally extended by Becker et al. [18] as “an event is a real-world occurrence e with a time period T_e and a stream of Twitter messages discussing the event during the period T_e”. Panagiotou et al. [81] describe that “in the context of online social networks, (significant) event e is something that causes a large number of actions in the online social network”.

Our study focuses on real-life events reported on social media. Therefore, we derive our definition as - “an event is a way of referring to an observable activity at a certain time and place that involves or affects a group of people in a social network”. The spatial and temporal coverage of a real-life event may vary depending upon its nature and intensity. A global event compared to a local event covers a broader scope and involves highly diverse content when reported on social media streams. Sometimes, global events comprise many local events. For example, an earthquake might be a global event involving many localities, and thus local events. Events are occurrent by nature, they take time to gain significance called “up time” and retain their importance and then end [23]. Accordingly, the information associated with an event can be categorized into three main phases: buildup, event-itself, and post-event effects [47].

Unlike a real-life event, a virtual event is defined as - “an event that takes place with participants who collaborate and interact without being physically present, connected by some form of technology” [14]. In addition to real-life and virtual events, trends on social media are topics that attract the attention of a large percentage of people such as #love, #food, and #happy which might not represent a real-life or virtual event. The focus of this article is real-life events such as a football match (sports), earthquake (natural), elections (politics), award ceremonies (showbiz).

2.2 Event Detection

People use online services to share content about various events they experience in their daily lives. Online communication services hold abundant and diverse contents shared by different people across the world regarding real-life events [48, 58, 60, 112]. With respect to social media content, event detection describes significant happenings in real-life by systematically analyzing the content published online and addresses how an event is emerging, gaining momentum, flows and evolves.

2.2.1 Specified Event Detection

When a social event is already known or planned, processing data concerning known information (such as location, time, keywords, and users) to extract event description is called specified event detection (SED). SED processes pre-defined information and the features which are expected to appear in the data to represent an event. This pre-defined information works as a seed to the actual event context [13].

2.2.2 Unspecified Event Detection

Unspecified event detection (UED) methods detect events in the absence of prior information. When social information about an event is unknown, a fundamental approach to UED is based on analyzing the temporal aspects of the Twitter stream by monitoring bursts to identify frequent keywords and concepts that are relevant to highlight events [13].

[3] http://plato.stanford.edu/entries/events/ (last accessed: January 28, 2019)

2.3 Supervised Learning

Supervised learning is an approach that generates class labels from training data. The training data consists of a set of examples (typically vectors) and a class label. Supervised learning analyzes the training set along with the class labels by producing an objective function. The objective function is then used to generalize the algorithm to classify other unseen data instances. The best-case scenario is that the class labels are correctly identified for all unseen instances. Testing and evaluation are mostly through error checks (i.e. correctly classified and miss-classified instances) [35].

2.4 Semi-supervised Learning

Semi-supervised learning lies between supervised (with completely labeled data) and unsupervised (with completely unlabeled data) learning approaches. Semi-supervised approaches process unlabeled data by exploiting partially available data. In such situations, acquiring labels involves human agents marking a small set of examples. In the context of the event detection process, this approach is also called a hybrid approach, in which the output is received by combining both supervised and unsupervised learning. One method is to use supervised learning in the filtration stage, and then unsupervised learning is used to group similar instances. The other method involves clustering similar instances first, then generating cluster labels through the objective function of supervised learning [98].

2.5 Unsupervised Learning

Unsupervised learning is an approach that summarizes or groups data instances (typically vectors) based on a similarity function. The given data is unlabeled. Therefore, there is no way of checking errors which distinguishes this approach from supervised learning. Testing and evaluation is carried out mostly by calculating inter/intra-cluster similarity [118].

2.6 New Event Detection (NED)

New event detection (NED) involves the continuous and live data monitoring of stream(s) of online/social content to detect the first trending story [67, 68]. In NED, the techniques are mostly developed around bursty features to detect a significant change in data using an unsupervised or semi-supervised learning approach.

2.7 Retrospective Event Detection (RED)

Retrospective event detection (RED) involves the task of detecting major events from historical data. The historical data can either be clustered or classified to detect significant events that happened in the past [54].

3 Data Collection and Evaluation

Twitter introduces several APIs for collecting data from its huge real-time repository. The purpose of providing open APIs is to promote external innovation. Offering information remotely through such open APIs permits researchers and developers to not only collect data easily but also to create innovative applications, platforms, and visual interfaces without the need to uncover crude information. Over 500 million tweets per day are published[4] on Twitter, and these tweets include lots of real-world information. In this section, we discuss different ways through which data can be collected from Twitter repositories and the datasets that have been used in recent studies.

3.1 Using Available APIs and Scraping

Twitter APIs allow data to be accessed programmatically. Two widely used public APIs for accessing Twitter data are *Streaming API*[5] and *Search API*.[6] These public APIs have certain limitations. In contrast with free APIs, paid services for Twitter data such as *Firehose*[7] and *Full Archive Search API*[8] are offered

[4] http://www.internetlivestats.com/twitter-statistics/ (accessed on January 28, 2019)

[5] https://developer.twitter.com/en/docs/tweets/filter-realtime/overview (accessed on January 28, 2019)

[6] https://developer.twitter.com/en/docs/tweets/search/overview (accessed on January 28, 2019)

[7] http://support.gnip.com/apis/firehose/overview.html (accessed on January 28, 2019)

[8] http://support.gnip.com/apis/search_full_archive_api/ (accessed on January 28, 2019)

Table 1 A comparison of four APIs available for accessing Twitter data

	Search API	Streaming API	Firehose	Full Archive Search API
Criteria	Involves polling Twitter data through a search criteria such as keywords, user names, locations, named places, etc.	User register criteria such as keywords, user names, locations, named places, etc. and as tweets match the criteria, they are delivered to the users.	Users register criteria such as keywords, user names, locations, places etc. Tweets are returned whenever the given criteria matches.	Crawl historical data through extensive search criteria such as keywords, user names, locations, named places etc.
Orientation	A basic historical search is performed. Tweets can be retrieved only for the past 7-15 days. It is more like a pull of data from history.	Works on real-time data, and whenever tweets appear that match the registered criteria, they are delivered as they happen.	Works on real-time data, data is delivered whenever the registered criteria is matched.	Provides filtered access to the full historical archive of public Twitter data.
Coverage	Comes with rate limits, against search requests it can pull approx. 3200 tweets and up to 5000 users.	Provides only a sample of tweets that are occurring. Studies show approx. 1% to 40% of total tweets can be collected [79].	Guarantees to deliver 100% of the tweets that match the given criteria.	Guarantees to deliver 100% of the tweets that match the given criteria back to 2006.
Limit	15 minute chunks per endpoint, Total of 180 search requests can be generated per-user-basis. Each user needs a token for access. There is a 100, 000 token limit for user accounts using the Twitter client app.	Limit for streaming API is on connection requests. Increasing number of attempts leads to rate limit, However, a few dozens attempts are allowed for connection for development purpose. Wait time for reconnection starts from one minute and doubles with every new request.	Connection attempts are limited to 10 requests per 60 seconds. Recommended connection wait time starting from one seconds and double with every new request	120 search requests per minute for an average of 2 queries per second. Data can be retrieved faster using multi-threaded approach. All requests from multiple threads are counted towards rate limit.

by GNIP[9] which guarantees to provide 100% of the tweets. The basic differences between the four APIs for accessing Twitter data are shown in Table 1.

For collecting data, most of the research studies use *Streaming API* [29, 34, 88, 90], or *Search API* [27, 59, 61], whereas limited research studies use *Firehose* [30, 54]. Access to Twitter Firehose is costly which might be one of the reasons it is not used commonly in research studies. In addition to the cost of Firehose, a recent research study, which collected 28 days of Twitter data with approximately 43% data coverage using Streaming API, shows that there is no significant difference between Streaming API and Firehose concerning data quality[79]. Similarly, topic discovery and network level measures are correlated to Firehose data even on temporal coverage as low

[9] http://support.gnip.com/apis/ (accessed on January 28, 2019)

as one day. Most interestingly, more than 90% of geo-tagged tweets coverage is seen in Streaming API. Thus, streaming API can be used as an alternative to Firehose, especially when the temporal coverage of the data is high.

In addition to the usage of *Search API*, which limits its temporal results to the past 7-15 days, some studies [51, 68] also used scraping for data collection to avoid this limitation. Scraping is employed by providing search keywords directly to the Twitter interface to get historical tweet. A web scraper parses HTML pages and grabs all the matched tweets automatically. However, an issue with web scraping is that it cannot be used to collect data in the context of NED. It is extremely challenging to gather live Twitter data using web scraping. The second major issue with this approach is that the Twitter server can block rapid HTTP requests for security reasons and the prior consent of Twitter is required before scraping the contents.[10] Scraping can also be used to collect data related to specified events by providing seed keywords [51]. In addition to the limitations of this data collection approach, scraping is time-expensive as a data collection strategy when high data coverage is essential.

3.2 Keyword-based Collection and Filtration

There are different ways of acquiring data, as discussed in Section 3.1. A keyword-based collection is a way to gather tweets which are related to certain events. It reduces the overhead of pre-processing with better data coverage [27, 59, 61, 88]. A keyword-based collection relies on seed-word(s) which represent topic(s) of an event. The data is acquired using seed-word(s) and all the tweets containing one or more seed-word(s) are collected. Keyword-based data collection is equally supported by both the Streaming and Search APIs.

A fraction of the collected tweets might not be related to any event. Therefore, the data is further contemplated through a filtration process which excludes irrelevant tweets. Due to the huge size of the data, filtration is mostly performed using automated processes by creating a lexicon from a source such as online news during the same period as the tweets [27, 118], a domain-specific description using the natural language processing [45], Wikipedia corpus [62], or by using a classifier to separate tweets representing events [66, 118]. Filtration is also performed for tweets that are less significant based on features like word count [115]. Other methods like crowd-sourcing and human labeling [27, 90] are also used, but due to the large volume of data, this is expensive and time-consuming.

3.3 Dataset Temporal Coverage and Size

The data is mostly collected for those events in which people have a high interest and participation level, whether local such as politics [88], festivals [61], rain [59] or global, such as World Cup [3], Show business [30, 62] and epidemics [59, 99]. For events like sports, politics, show business, disease outbreak, and natural events, abundant data is available. An increasing number of people interested in an event increases the chances of getting high coverage on the Twitter stream. Still, obtaining quality data for a popular event cannot be guaranteed.

There are different services that offer open data sets to research communities such as SNAP,[11] TREC[12] and GNIP. Most researchers use Twitter APIs to collect data directly from Twitter. A few studies use existing datasets, especially when comparing and evaluating different techniques. One significant aspect about the available datasets is that they do not contain content but user and tweet IDs only. Distributing content, whether user profiles or tweets, is controlled by legal privacy restrictions defined by Twitter and is not permitted. Therefore, the available datasets only contain IDs. One must trace and crawl the content of the IDs through Twitter APIs.

Using public Twitter APIs (i.e. Streaming API and Search API) is the most common way of crawling tweets for generating datasets. However, some of the studies also collected data from paid data services such as Firehose and Gardenhose. *Scraping*, an infrequent and complex method, has also been used in a few research studies. The temporal coverage of data varies significantly from a few days to one year. The average temporal coverage of datasets

[10] See Section 4 Using the Services @ https://twitter.com/en/tos (accessed on January 28, 2019)

[11] https://snap.stanford.edu/data/index.html (accessed on January 28, 2019)

[12] http://trec.nist.gov/data/microblog.html (accessed on January 28, 2019)

is six months. Similarly, data size also varies significantly from a few thousand tweets to hundreds of millions of tweets. For an approximate estimation regarding datasets and their coverage, we calculated average dataset sizes from the various studies reported in our survey for specified and unspecified categories. On average, each study uses 26 million tweets if we ignore the extremely large or small datasets with 3 billion tweets [38], 442 million tweets [30], 10 tweet [83] and 597 tweets [93], 1.16GB data [119], 400GB data [45] and Firehose datasets [84, 85] where data size is not mentioned. Interestingly, studies which examine unspecified events consider 30.8 million tweets on average, i.e., 45% greater than the average of 21.1 million tweets for specified event detection studies. The reason for the smaller data size for specified event detection is that studies usually narrow down the scope with seed keywords related to targeted events and acquire data that match the criteria. In the case of unspecified event detection, event-related information is unknown, therefore acquired data may contain heterogeneous events and as well as greater noise; hence greater data coverage is essential to detect meaningful patterns.

A detailed comparison of data collection strategies is discussed in Section 3.1. Further details on the datasets are summarized in Table 2 for unspecified and Table 3 for specified events.

3.4 Evaluation Methods

Evaluating an event detection technique is essentially an important task. Typically, event detection methods produce a ranked list of keywords. The output keywords are considered as key concepts representing an event. Finally, The output keywords matching with the ground-truth are empirically quantified to evaluate the performance of an event detection technique. The methods to create benchmark dataset for ground-truth and evaluation measures are described in the following Sections 3.4.1 and 3.4.2 respectively.

3.4.1 Benchmarking Dataset

Evaluating a technique is a challenging task in the absence of a benchmark dataset. Benchmarking is a difficult task as there are a variety of events which are diverse in many ways such as, popularity, user participation, and content size. A standardized benchmark that can cope with diverse events related to different domains is very costly, challenging and unavailable to date to the best of our knowledge. This limitation can be addressed through the filtration process. Filtration is the initial step to reduce data size using criteria based upon tweet size, language, users and bursty features and then ground-truth can be created with minimal cost.

In the literature, we find three different approaches to create benchmark dataset. The first approach considers the events reported in mainstream news in a period similar to the data collection and uses them as ground-truth [4, 27, 118]. This approach can be useful when analyzing major or out-breaking events. However, small scale events such as conferences, local festivals and tweet rumors are not always reported through mainstream news media, hence might be ignored. The second approach uses manual labeling. Ground-truth is created by selecting *N* random tweets from the dataset that can easily be labeled by humans [54, 90]. The third approach uses clustering algorithm to segregate the tweets into clusters and then automatically labels these clusters through bursty features [45, 51]. The results are then manually observed and evaluated by human experts [115, 118] and sometimes external cluster evaluation is used as a gold standard [51]. The labeled data then serves for performance evaluation for the event detection techniques. However, aforementioned methods for creation of benchmark datasets are not mutually exclusive and sometimes used in combinations [118].

3.4.2 Evaluation Measures

To evaluate the results, *Precision*, *Recall*, and *F1-Measure* are the most commonly used evaluation measures. Similarly, *Accuracy and Error Rate* are also among frequently used measures. Some studies also use *Qualitative* [2, 15, 61, 73] evaluation.

Despite the limitations of *Accuracy* as an evaluation measure, we found that it is widely used in evaluating event detection methods [10, 30, 38, 46, 54, 66, 103]. It evaluates performance of the classifier as *Accuracy = (TP+TN)/(TP+FP+TN+FN)*.[13] It is acknowledged that high accuracy may not be an indicator of the better performance of a classifier. Accuracy is sometimes

[13]TP=True Positives, FN=True Negatives, TN=True Negatives, FN=False Negatives

Table 2 Dataset details and evaluation measures for unspecified events

Ref	Collection	Corpus size	Temporal Scope	Evaluation
Abdelhaq et al. [2]	Streaming API, Search API	Live stream	Continuous	Qualitative
Alsaedi et al. [10]	Streaming API	1.6M tweets	20 days	Precision, recall, F1-measure, and accuracy with 10-fold cross-validation
Becker et al. [18]	Streaming API	2.6M tweets from NYC-based users	1 month	Macro-averaged F1-measure
Chen et al. [27]	Search API and crawler for online news	51K, 130K, and 142K tweets from Singapore based users and companies	— —	Precision, recall, F1-measure and qualitative
Cheng and Wicks [29]	Streaming API	1.85M tweets	12 days	— —
Cordeiro [31]	Spritzer	13.6M tweets	10 days	Visual illustration
Cui et al. [34]	Stanford SNAP data and Streaming API	2M and 778K tweets respectively	6 months and 3 months respectively	Precision, recall and F1-measure
Gao et al. [38]	Sina Weibo API	3B micro-blogs and 170k users from Seina Weibo	2 months	Accuracy
Huang et al. [46]	Search API, Sina Weibo API	2M tweets and 280K micro-blogs from Sina Weibo	6 months & 6 months respectively	RI (Rand-Index), TPR (true + rate), FPR (false+ rate) and F1-measure
Li et al. [62]	— —	4.3M tweets and 3.2M articles and 266M hyperlinks from Wiki dumps	1 month and Wiki dumps from 30-Jun-10	Precision, recall, and Wikipedia dumps to create newsworthy segments
Liu et al. [67]	Sina Weibo API	2M micro-blogs	17 days	Precision, recall and F-score
Long et al. [70]	Sina API	22M posts	2.5 months	Precision
Petrović et al. [82]	Streaming API	163.5M tweets	6 months	Average precision
Phuvipadawat and Murata [83]	Streaming API (predefined search queries targeting breaking news)	10 Tweets	— —	Qualitative
Ritter et al. [90]	Twitter streaming API	100M recent tweets	3rd Nov 2011	Precision, Recall and F1-measure
Sankaranarayanan et al. [94]	BirdDog	— —	— —	Qualitative
Tu and Ding [103]	Sina Weibo API	— —	1 month	Accuracy
Weng and Lee [112]	REST API	4.3M tweets	1 month	Precision

Table 2 (continued)

Ref	Collection	Corpus size	Temporal Scope	Evaluation
Zhang et al. [115]	Streaming API and Sina Weibo API	31M tweets of 313K randomly selected users, and 6M micro-blogs (Sina Weibo) of 119K users from the recommended topic	1 year and 6 months respectively	MacroPrecision@K
Zhou et al. [118]	Streaming API	60M tweets	1 month	Precision, recall and F1-measure
Zhou and Chen [119]	— —	53.4M and 1.16G tweets	3 days and a month respectively	Probabilities of missed detection and false alarm errors

misleading in the case of asymmetric class distribution, which is also called the accuracy paradox [106, 120]. Consider the example given in Table 4. According to the accuracy measure, the better performance is 98.5% which is misleading because in this case none of the TPs (True Positives) are identified. So, whenever TPs $\ll$ TNs, then accuracy will always increase when the classification rule is changed to always output the "negative" class. Conversely, when TP $\gg$ TN, the same will happen when the classification rule is changed to always output the "positive" class. In the context of event detection on Twitter, TPs are more important, and in fact, the performance of a technique relies on identifying TPs. Secondly, approximately 57% of tweets are noise or irrelevant and belong to the "negative" class. In addition, 38% are conversational tweets that may or may not belong to an event (See Section 7 for details). The output rule, especially in the filtration process to reduce irrelevant content is set to output negative class, in which case, this phenomenon is more likely to occur. Therefore, accuracy is misleading in the presence of this paradox. It is difficult to tell if a study has encountered the accuracy paradox. None of the studies used in our survey highlighted this issue. However, there is a chance that this issue might have arisen in some studies. Thus, before evaluating the performance of a system, one should look at the class distribution, and use an appropriate evaluation measure.

To avoid the paradoxical phenomenon in accuracy, *precision* and *recall* along with the *F-measure* can be used. In the case of using recall as an evaluation measure, it is necessary to know all the real-life events that exist in the data which is very difficult as data labeling is expensive and mostly unavailable. Therefore, sometimes only precision is used especially when event-related information is unknown [112, 115, 118]. Some studies, after screening out irrelevant tweets and reducing the data size, manually label the data to undertake the performance evaluation and use precision, recall, and F-measure collectively [34, 118]. In the case of ranked results in an evaluation, where the top ranked results are more relevant, *Average Precision* and *Precision@K* are better measures to use [10, 76, 82, 85, 115]. If events are critical such as disasters, emergencies, crimes, or security-related, then it is important to detect events at an early stage from the data stream and it would be useful to measure the time taken by the event detection technique for generating ranked results.

In addition to quantitative evaluation, qualitative evaluation is also used in many recent studies [14, 27, 73, 83]. This involves human experts validating the results in the absence of labeled datasets. A summary of the evaluation measures can be seen in Table 2 for unspecified and Table 3 for specified events.

4 Feature Extraction

The richness of Twitter data (see Section 1) provides a great opportunity to explore new and existing features to be used in the event detection process. There are seven aspects that researchers consider while studying micro-blogging systems [113], of which *message*

Table 3 Dataset details and evaluation measures for specified events

Ref	Collection	Corpus size	Temporal Scope	Evaluation
Adedoyin-Olowe et al. [3]	Search API	224k, 3.8M, and 474K tweets	2 hours, 1.5 days, and 1 day for each tweet set respectively	Support, confidence, precision, recall and F1-measure
Becker et al. [15]	Twitter API	——	——	Qualitative
Benson et al. [20]	Public API	2.7M tweets	3 weekends	Precision, Recall
Chen et al. [28]	Sina Weibo API	22.3M micro-blogs	1 month	Precision, Recall
Chierichetti et al. [30]	Firehose	442M, 1.49M and 1.61M tweets	1 month, 1 day and 1 day respectively	Error rate (false+, false-, true+, true-)
Gu et al. [41]	Search API	3.5M tweets	5 months	Coverage/coherence
Hua et al. [45]	——	400GB in size	8 months	Precision, Recall and F1-measure
Kaleel and Abhari [51]	Scraping	1M tweets	4 days	Group-Average Agglomerative Clustering (GAAC) as gold standard and purity, normalized mutual information as external criteria
Khurdiya et al. [54]	Garden-hoses	13.32M tweets	——	Precision, Recall and F1-Measure
Lampos and Cristianini [59]	Search API	8.5 M and 50M geo-tagged tweets	12 months and 10 months respectively	6-fold cross-validation and linear correlation
Lee and Sumiya [61]	Search API	21.6M geo-tagged tweets	1.5 months	Precision and recall
Li et al. [65]	Sina Weibo API	3M microblogs and 100K users	10 days	Average precision and recall
Li et al. [66]	Search API	1M tweets	2 months	Accuracy
Massoudi et al. [73]	Search API	110M tweets	10 months	Average Precision, Mean Average Precision
Metzler et al. [76]	Twitter Streaming API	46.6M tweets	6 months	Precision@K
Popescu et al. [85]	Firehose	5040 snapshots	——	Average precision and mean reciprocal rank
Popescu and Pennacchiotti [84]	Firehose	740K tweets	7 months	Receiver operating characteristics, average precision
Rill et al. [88]	Twitter streaming API	4M tweets	5 months	Correlation with Google trends
Sakaki et al. [93]	Search API	597 tweets	——	Precision and recall

and *user* are the most commonly used. In the context of event detection on Twitter, the messages (tweets) are interpreted and manipulated to extract meaningful and useful information. Classic IR features such as *tf* and *idf* serve as key features to begin with. New features which satisfy the need for event detection can be extracted. There are various custom features, but we discuss only those which can be used across different datasets and techniques for event detection in Sections 4.5 and 4.4.

Table 4 Example showing the accuracy paradox. The result on the right side has higher accuracy, but is misleading

Accuracy of 98%			Accuracy of 98.5%		
	Predicted (YES)	Predicted (NO)		Predicted (YES)	Predicted (NO)
Actual (YES)	100	50	Actual (YES)	0	100
Actual (NO)	150	9700	Actual (NO)	50	9850

4.1 Keyword-Based

The term vector (including *tf* and *idf*) is the most useful feature, and it can be extracted directly from tweets [27, 29, 30, 45, 51, 119]. Similarly, co-occurrence frequency of a set of words is also useful to identify a group of bursty words [4, 46, 115]. However, while considering the dynamics of event content and user sentiments, these features are used along with natural language-based features. The language-based features highlight the significance of the keywords used in event content which cannot be captured by keyword-based (bursty) features. Several other studies [4, 39, 59, 62, 66, 98] prefer using N-grams by justifying that bursty segments are more event expressive than single words.

4.2 Twitter-Based

Hashtags, user information, time-stamps, retweets, and geo-tags are among the Twitter-based features used in various studies. Hashtags are considered explicit content descriptors and frequently appear in event contents [34]. It is observed that approaches which typically follow clustering techniques do not consider hashtags to be a distinct feature rather they are used in the bag-of-words model [76, 115]. Most of the event detection approaches for unspecified events do not target specific keywords in the data. Therefore, they give equal weight to hashtags that are an integral part of tweets.

People start retweeting when an event occurs [30]. Retweeting adds more frequency burst to keywords that are actively used to report an event, so a *retweet* is another feature which is common. However, despite a retweet's importance, it adds redundancy into the Twitter text stream and raises scalability issues as discussed in Section 7. Despite this redundancy, retweets are useful and are used in many studies [1, 27, 30, 45, 54, 66]. Retweet feature is sometimes misleading and induces bias in the bursty features. Techniques for event detection must normalize the retweet effect in conjunction with user participation. Otherwise, the bias factor of influential users may affect the outcome.

4.3 Location-Based

Increasing use of smart devices (smartphones, tablets, handheld digital assistants) enables geo-tagging of tweets seamlessly. Geo-tagging is one of the important features which is widely used by research studies and plays an important role in spatial event detection. Geo-tags are not available in a significant number of tweets, with approximately only 2% of the total tweets are geo-tagged [45]. Acquiring geo-labels is a difficult task, but can be inferred by using k-nearest neighbor (KNN) tweets based on their locality [61] or by location mentioned in the tweets [9, 10, 90]. In addition to its usefulness, extracting the geo-location using KNN is helpful for small-scale local events, but the method for geo-label generation, such as described in [61], might not work if the event has a broader perspective or took place in several locations such as an election campaign. The location may mislead in the detection of several separate events instead of a single event. Nevertheless, extracting geo-locations from tweet content is useful for geo-sensitive events.

4.4 Language-Based

Features such as nouns, verbs and part-of-speech (POS) tags, are also important [9, 10, 103]. These features are more authoritative in terms of describing and expressing event-related information. Event information consists of three major parts, i.e., purpose (tweet content), time, and location and all the three dimensions (textual, spatial and temporal) are equally important when defining a machine learning technique for event detection. Recent studies [9, 22, 116] show that these dimensions do not give optimum results when used in isolation. Rather, it is best if all three feature dimensions are used in combination.

4.5 Custom

Handling user bias is a challenging task, as bursty features such as keywords frequency, retweets counts, and hashtags are sometimes misleading due to factors such as spam and advertisements. In addition to the well-known tf-idf that balances biasness, different features such as *user authority* [115] (for information diffusion), *hashtag instability* (for topic dispersion), and *authorship entropy* [34] (for popularity) could be useful when used in conjunction with bursty features to identify event significance. Another feature named *binary word* [118] is also useful. A binary word shows whether the underlying word is related to a specific event or not. The algorithm does so by classifying it using an event-related control vocabulary. Since it involves a domain vocabulary and classification, a binary word can only be used in a limited context for specified event detection using supervised approaches.

Selecting and using appropriate features is one of the crucial parts in the implementation of a machine learning technique. There is a rich set of features associated with Twitter contents. Therefore, it is crucial to select appropriate features. The features discussed above, are mostly used together or in combination. All of them have their unique significance. Therefore, depending on the technique and characteristics of the event being focused, these features are helpful in the event detection process. Tables 7 and 8 summarize the details on techniques and features used in various studies.

5 Event Detection in Twitter: Methods and Techniques

Twitter publishes a high volume of user-generated content. These content might represent real-life events and become a potential source of event information as shown in Fig. 1. "The event-related data falls into three major phases: the buildup to the event, the event itself, and the post-event effects and repercussions" [47]. The popularity and dynamic nature of Twitter allows us to record events in real-time. In many cases, the tweets are created by users who are either participating in an event or immediately affected by the event. For events such as the occurrence of a hurricane [66], a flood [119], an influenza-like illness [59], or earthquake [93], this data can be used to detect and analyze how the event progressed, and traveled geographically. The reactions of people to disaster relief efforts can also be studied and analyzed to make improvements in the future [47].

Despite the user-generated event information, there are many tweets generated to spread rumors and spams to take advantage of popular and out-breaking events. Approximately 0.13% of advertisement messages contain clickable URLs which are twice as higher than email spams. Due to the limit on tweet length, additional information is usually provided in the form of tiny URLs which makes Twitter an attractive social media platform for spammers [26, 95].

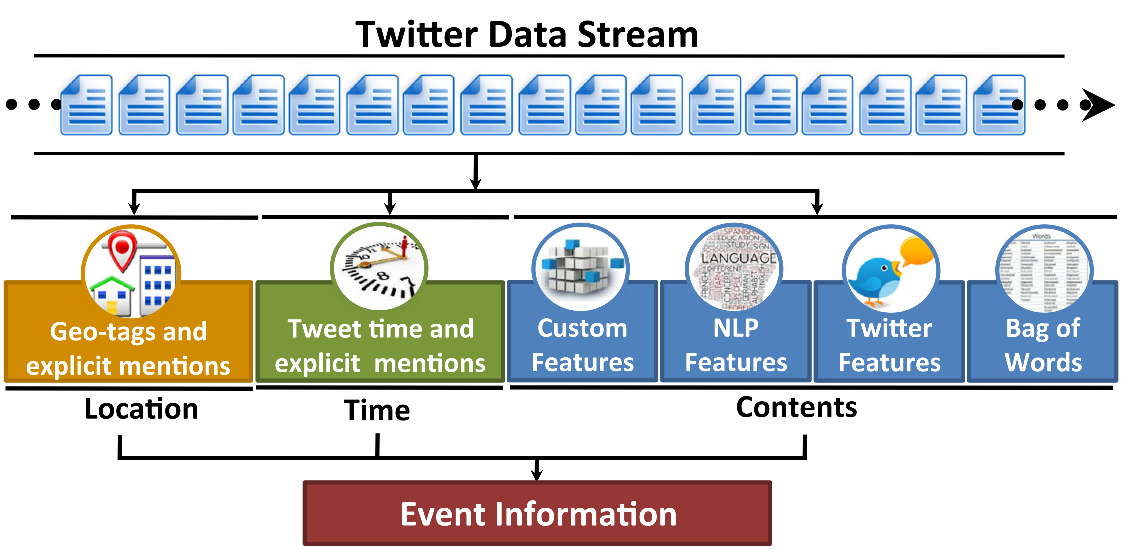

Fig. 1 Content features and event information on Twitter

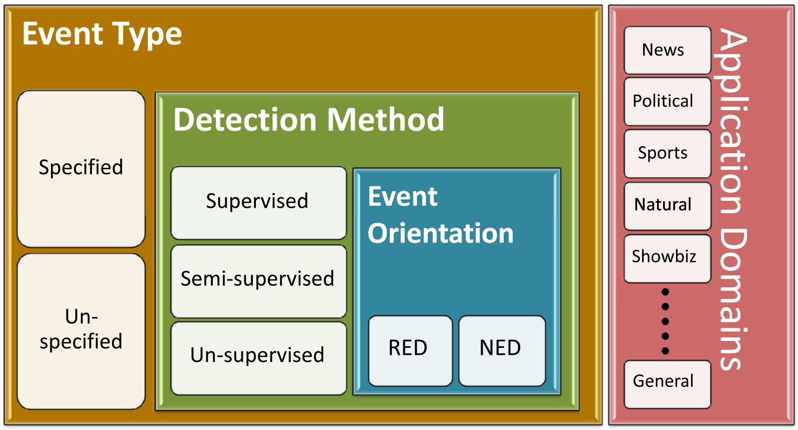

Fig. 2 Taxonomical hierarchy based on taxonomy proposed by Atefeh et al. [13]

On the other hand, a rumor is a statement whose truth value is unverified or intentionally false [68, 109]. A bot is an account that is controlled by computer programs. Activities of propagating spam content and rumors are usually carried out through bots [25]. These bots use trending keywords/hashtags to post tweets automatically with the desired goals such as advertising services, products, and participating in commercially sponsored campaigns. Detecting spams and rumors on Twitter is a well-researched topic. However, the detection techniques used in the research areas mentioned above is significantly different than the ones used in real-life event detection hence outreaching the scope of our current survey study. Therefore, we do not discuss such techniques in this paper.

To better understand various techniques proposed for event detection on Twitter, we extend an event detection taxonomy orthogonal to the one proposed by Atefeh et al. [13]. The taxonomy is created according to the type of event, detection approaches, orientation, and application domains. The taxonomy as shown in Fig. 2 has four main categories i.e. 1) event type 2) detection approach 3) orientation and 4) application domains. "event type" is further divided into two sub-categories: specified and unspecified. Studies using some pre-defined event information or known events are classified as specified, whereas studies which do not consider prior event information and mainly rely on bursty features to detect unknown events from Twitter data stream are categorized as unspecified. According to the detection models used in the existing literature, detection approaches are further divided into three sub-categories: supervised, semi-supervised and unsupervised. Studies which use hybrid (i.e., combination of classification and clustering) approaches fall into the semi-supervised category. Orientation is further divided into new event detection (NED) and retrospective event detection (RED). The techniques which utilize a live Twitter stream to detect new and emerging events are classified as NED, and those which use historical data are classified as RED. The taxonomic hierarchy is visually presented in Fig. 2. Furthermore, we also classify studies with respect to their social context and application domain such as politics, nature, sports, and general or unknown. The application domain is listed in alphabetical order, and is independent of the hierarchical categories (i.e., Event Type → Event Detection Method → Event Orientation). We have added some of the domains which are commonly used as case studies. This classification gives an insight into event detection approaches and the importance of the different events in which researchers are taking interest and currently working on. To avoid sparseness in the tabular structure and to obtain an overview of the categorical details of the studies, we created a two-dimensional index of techniques (listed in Table 6) for all the studies. This index is used in Fig. 3 and Table 5.

The research papers belonging to various categories in the hierarchy are listed in Fig. 3. Research papers belonging to a specific category are listed at the leaf nodes of the taxonomy. For example, studies $F2$ and $C1$ which belong to $Specified \rightarrow Semi-supervised \rightarrow NED$ and $Unspecified \rightarrow$

Table 5 The classification of studies with respect to application domains

Application Domains	Indexed References of Studies
News	D3, D4, D6
Sports	B6, D5, E2,
Political	D5, E3,
Crime/Unrest	E5, F2,
Natural	A5, F1, F2, G5
Epidemic	F1
Conference	B6
Showbiz	B6, F4, F5, G4
General Unknown	A1, A2, A3, A4, A5, A6, B1, B2, B3, B4, B5, B6, C1, C2, C3, C4, C5, C6, D1, D2, E1, E6
Query-based	B6, E3, E4, E5, F2, F3, F6, G1, G2

The second column represents the indexes of the existing studies given in Table 6

$Supervised \rightarrow RED$ categories respectively. There is no study in the $Unspecified \rightarrow Supervised \rightarrow NED$ category because supervised approaches need class labels for training and for unspecified events these labels are missing. On the other hand, most of the techniques belong to the categories: $Specified \rightarrow Unsupervised \rightarrow RED$, $Unspecified \rightarrow Unsupervised \rightarrow NED$, and $Unspecified \rightarrow Unsupervised \rightarrow RED$ (Fig. 3).

Table 5 provides an overview of the techniques applied in specific domains. We can observe that most of the papers detect generic or unknown events and there are only two papers which investigate epidemics and conference events. Event detection techniques and features used by these techniques are summarized in Table 7 for unspecified and in Table 8 for specified events.

Details on different research studies and their techniques are given in the following Sections 5.1 and 5.2.

5.1 Specified Event Detection (SED)

A variety of text analysis and machine learning techniques are designed to analyze Twitter data stream [13]. The following sections describe the detail about detection approaches for SED.

5.1.1 Supervised Approaches

Khurdiya et al. suggest that big events such as the Academy Awards and Football World Cup, which have comparatively high user participation, are normally planned and advertised in advance, but smaller

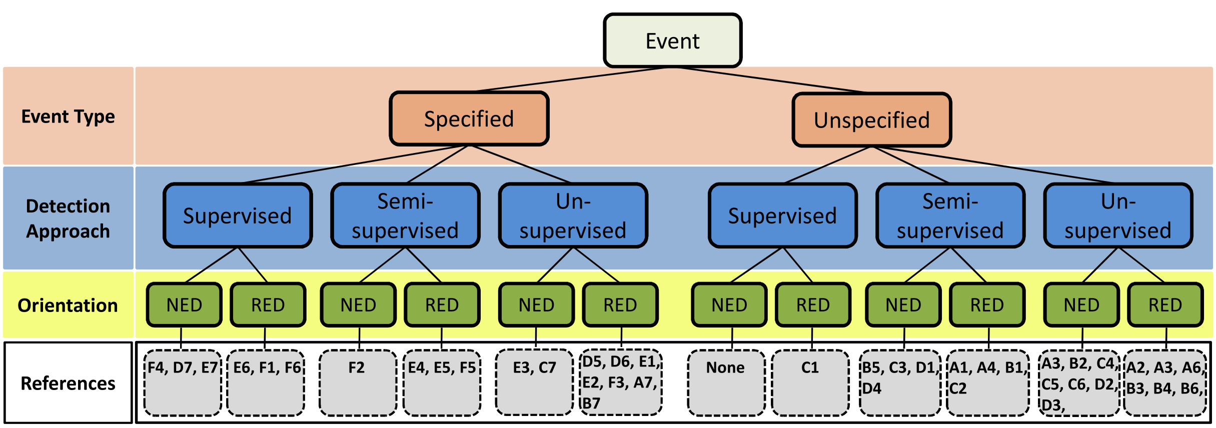

Fig. 3 Visual taxonomy of event detection techniques in Twitter. Each dotted leaf node is a group of indexes representing papers that are categorized under the hierarchy. The papers against the indexes are given in Table 6

Table 6 Indexes of all the studies used in this survey

	A	B	C	D	E	F
1	[46]	[10]	[34]	[18] *	[65]	[59]
2	[115]	[2]	[103]	[82] *	[30]	[66]
3	[118]	[38]	[90]	[83] *	[88]	[76] *
4	[9]	[67]	[31] *	[94] *	[28]	[85] *
5	[119]	[27]	[70] *	[3]	[45]	[20] *
6	[29]	[62]	[112] *	[51]	[54]	[15] *
7	[73] *	[41] *	[61]	[84] *	[93] *	

Details of studies with "*" can be seen in [13]

events such as protests and threats which are built around major events are not known in advance [54]. The focus of their study is "to identify, extract and build a map of small sub-events around a big popular event" [54]. They also developed a framework which uses *Searching on Lucene with Replication* (SOLR) and *Conditional Random Field* (CRF) for event extraction and title identification respectively. They also proposed that "along with an activity, an event can be additionally associated to a combination of elements like subject, object, and context of the activity. A single event is spread over time and location, whereas the sub-events have a combination of elements that remain fixed while others can change". Six components are proposed to define an event i.e. subject, action, object, time, location and additional description (contextual information) which serves the purpose of identifying the sub-events emerging around some big event. To evaluate their proposed model, 3000 random tweets are selected from the dataset and manually labeled. Then, 2-fold cross-validation is used to evaluate the performance.

Lampos et al. propose a statistical learning framework to identify an event by mining a huge volume of textual information [59]. They use "Least Absolute Shrinkage and Selection Operator" (LASSO) for feature subset selection. The methodology consists of the following three operations: 1) candidate feature extraction: N-grams vocabulary of features is formed 2) vector space representation (VSR): for a fixed time period and set of locations, the VSR of the candidate features are computed from term frequencies (TF) 3) feature selection and inference: sparse regression method is used to select a subset of candidate features. Candidate features are extracted from the websites of the National Health Service, BBC, Wikipedia, and weather-related websites. Experiments were conducted on two different datasets for specified events: one for rainfall rates in five urban cities of the UK and the second for flu rates in three regions of the UK. Their results for rainfall rates show an improvement of more than 10% by using both Hybrid 1-grams and 2-grams. The results are compared with other baseline approaches using 5-fold cross-validation. The results for the second case study, the diffusion of influenza-like illness, shows flu rates reach an average correlation of 91.11% with the actual ones.

5.1.2 Unsupervised Approaches

Frequent pattern mining [64, 97] is one of the many unsupervised approaches to capture word co-occurrence and is used to detect specified events. The study [3] considers hashtags that evolve over time as a significant and primary feature to define rules. To achieve this, it takes three events i.e., the FA Cup Final 2012, the US Election 2012 and the Super Tuesday 2012 for the collection of tweets based on related keywords. Each event is divided into smaller chunks into time slots of 1 minute, 10 minutes and 1 hour, based on the evolving rate of each event. The support and confidence are set to 0.001, which despite being low, allows abundant itemsets of hashtags related to the event to be extracted. Hashtags that meet minimum support are ranked on three levels, i.e., unexpected consequent/conditional rules, emerging rules and a combination of both. The hashtags returned by the association rules are then matched with the ground-truth. Event detection occurs when the time-slot has

Table 7 Summary of event detection techniques and feature representations for unspecified events

Ref	Detection Techniques	General/Extracted Features	Twitter-based Features
Abdelhaq et al. [2]	Single-pass clustering	Temporal word frequency	Geo-tags, time stamp
Alsaedi et al. [10]	Naive Bayes classifier and online incremental clustering	Term vector, tf-idf, location, unigram, bigram, POS, named entities	Time stamp, geo-tags, hashtags, mentions, URL, retweet
Becker et al. [18]	SVM and online clustering	Term Vector	Hashtags, multi-word hashtags, retweets, replies, mentions
Chen et al. [27]	Single-pass Incremental clustering	Term vector	Users, tweet count, retweet, count, followers
Cheng and Wicks [29]	Latent dirichlet allocation (LDA), space-time scan statistics(STSS), space-time permutation model (STPM)	Term vector	Time stamp, geo-tag
Cordeiro [31]	LDA and Continuous wavelet analysis	— —	Hashtag occurrence
Cui et al. [34]	Subspace-based algorithm	Hashtag instability, meme possibility, authorship entropy	Hashtags, users
Gao et al. [38]	Adaptive K-mean clustering	Term vector, tf-idf	Geo-tags, time stamp
Huang et al. [46]	K-nearest neighbor and modularity clustering	Words frequency, co-occurrence, POS and support	— —
Li et al. [62]	Jarvis-Patrick algorithm	Tweet segments	Users
Liu et al. [67]	Single-pass clustering	Unigram noun and verbs	Time stamp
Long et al. [70]	Hierarchical clustering (divisive)	Entropy	hashtags
Petrović et al. [82]	Online clustering (based on locality sensitive hashing)	Entropy	Tweet count, users
Phuvipadawat and Murata [83]	Online Clustering	Term vector, proper nouns, named entities	Time, retweets, hashtags, follower count
Ritter et al. [90]	Conditional Random Field (CRF)	Named entities, POS tags, location	— —
Sankaranarayanan et al. [94]	Naive Bayes classifier and online clustering	Term vector	Time, hashtags
Tu and Ding [103]	Bayes classification, mean calculation-based top k-category incomplete clustering algorithm	Noun, verb, tf-idf and word length	— —
Weng and Lee [112]	Discrete wavelet analysis and graph partitioning	Term frequency	— —

Table 7 (continued)

Ref	Detection Techniques	General/Extracted Features	Twitter-based Features
Zhang et al. [115]	Graph Cluster Algorithm to find strongly connected components using depth-first-search technique	Term vector	Time stamp, users
Zhou and Chen [119]	Location-time constrained topic (LTT) model and similarity query algorithm	Term vector	Time stamp, user, geo-tags, followers
Zhou et al. [118]	Latent Event & Category model	Binary word, opinionated words, selected phrases, named entities, POS tags, currency & percentage sign, time, location	Time stamp, URL, mentions

at least one keyword similar to ground-truth in the same time frame. To evaluate this technique, ground-truth is generated from mainstream media for the political dataset and the BBC official website for the sports dataset. The study concludes that hashtags and emerging rules represent event highlights.

Kaleel and Abhari introduce the Locality Sensitive Hashing (LSH) technique for event detection and trending [51]. The technique is based on classic IR features with a novel indexing mechanism. As an initial step, only qualified tweets are taken with the criterion of having at least 30% of the words in English. Tweets are then tokenized and stop words are removed. URLs and mentions are also removed for simplicity. Due to the huge data size, updating the term vector and changing the dimension of the vector when a new word arrives is challenging. A combined approach is used for updating the term vector with Incremental tf-idf. Instead of a single dictionary, chunks of dictionaries were created where N tweets are indexed in each dictionary. After reaching capacity, a new dictionary is created each time. A high dimensional vector is converted into a K-bit signature while preserving the cosine similarity among the term vectors. These K-bit signature features from all dictionary chunks are then used for the clustering with cosine similarity. Most frequent terms within a cluster are used for the centroid and label the cluster. Manual labeling is expensive when the data size is huge. Therefore, group-average agglomerative clustering (GAAC) is used as the gold standard. The quality of clusters discovered by K-means and LHS are then measured against GAAC with purity and normalized mutual information (NMI) as external evaluation criteria. The results show that LSH is 12.5% and 16.6% better than K-means in purity and NMI respectively. In this study, the bias of redundant tweets and retweets is not considered. Therefore, tweets from influential users [53] may affect the quality of cluster labels. User influence is helpful when predicting events as in [115], but can be misleading for event detection at the same time.

In the presence of linguistic issues such as multiple languages, language diversity and heterogeneity, abbreviations and grammatical errors, it is a challenging task to process noisy textual content for event detection. Due to linguistic issues, Chierichetti et al. in their study [30] focused on the non-textual features of tweeting patterns, such as time and retweets, to detect important events. In addition to event detection, they identified how an event affects users' communication behavior in terms of producing new information and forwarding existing information by concluding that users often reduce the volume of communication with other users (i.e., mentions, replies) and increase their retweeting activity regarding an external event that happens at a specific time. Their model classifies tweets into pre-defined categories using an unsupervised classification model and assigns each user to one of the 32 teams participating in the football World Cup using geographical, language information, and hashtag usage. A user who is assigned to a team is considered to be a supporter of that team. They found that tweet volume is skewed towards the team winning

Table 8 Summary of event detection techniques and feature representations for specified events

Ref	Detection Techniques	General/Extracted Features	Twitter-based Features
Adedoyin-Olowe et al. [3]	Association rule mining	— —	Hashtags
Becker et al. [15]	Recursive query construction	Location, term frequency	Hashtags, URLs
Benson et al. [20]	CRF and factor graph model	Term vectors, word shape, emoticons	Time
Chen et al. [28]	Logistic regression classifier, and Affinity propagation clustering	Term vector, tf-idf	— —
Chierichetti et al. [30]	Logistics regression approach	Word frequency	Time stamp, retweets, replies, users, hashtags
Gu et al. [41]	Events modeling	NGrams	Replies
Hua et al. [45]	Graph partitioning and SVM	Noun,verbs, term vector	Retweet, mentions, hashtags, geo-tags
Kaleel and Abhari [51]	Locality Sensitive Hashing (LSH)	Term vector	Users, geo-tags
Khurdiya et al. [54]	SOLR and CRF-based event extractor	Root words, capitalization, POS tags, named entity, tags (people,location,date)	Hashtags, user mentions, re-tweets, time, geo-tags
Lampos and Cristianini [59]	Bolasso-S (for feature selection) and ordinary least squares (OLS) regression	Term vector, 1-grams (U), 2-grams (B), hybrid (H)	Geo-tags
Lee and Sumiya [61]	K-means for Region of Interests and statistical model for detecting unusual user behavior	moveCrowd	Tweet count, users, geo-tags
Li et al. [66]	Naive Bayes classification model	1-grams and 2-grams	Hashtags, mentions, retweet count, favorite count, users
Massoudi et al. [73]	Generative language modeling	Tweet length, capital letters, emoticons	URLs, time, retweets, follower count
Metzler et al. [76]	Temporal query expansion technique	Query keyword frequency	— —
Popescu et al. [85]	Gradient boosted decision trees	POS tags, regular exp. position information	— —
Popescu and Pennacchiotti [84]	Gradient boosted decision trees	Correlation of target events, entities with the Web & traditional news media, nouns, verbs, bad words, questions, sentimental and controversy words	Tweet count, retweet count, replies, hashtags

Table 8 (continued)

Ref	Detection Techniques	General/Extracted Features	Twitter-based Features
Rill et al. [88]	Standard deviation in candidate topic	——	Hashtags
Sakaki et al. [93]	Suport vector machine (SVM)	Query terms and keywords	——

the ongoing match and users communication patterns show the occurrence of the event. They concluded that communication patterns can contribute very well to the detection of major events on Twitter.

Rill et al. studied the identification of emerging political events [88]. In 2013, during the German elections, Twitter data was collected for their case study, but the focus was not to forecast the election results but to detect an early emerging political topic. For topic detection, hashtags are considered as candidates for the top topic. Their technique finds a standard deviation distribution of frequencies for candidate topics considering the time window. For topic classification, as top topics, topic value (tv) is calculated and normalized from -1 to 1 where $tv > 0$ indicates the topic is gaining interest and emerging.

Lee and Sumiya present a model to detect geo-social event by utilizing the collective experiences and behaviors of crowds over Twitter [61]. They developed a geo-social monitoring system consisting of the following modules: 1) collecting crowd experiences (tweets with geographical locations); 2) setting out the regions of interest (RoIs) using k-means clustering which later forms a Voronoi diagram using centers of clusters; 3) estimating the geographical regularities of each RoI based on user activities in four equal temporal windows of a day. Geographical regularities are calculated by considering the questions such as *"How many tweets are posted?"*, *"How many users are there?"* and *"How active are the movements of the local crowd?"*. These observations on historical data regarding RoIs define geographical regularity as a normal tendency; 4) detecting unusual crowd activities by using geographical regularity as an indicator. Crowd activities are classified as high and low significance by comparing them with geographical regularities. RoIs with high significance are considered to be part of the geo-social events. The authors conclude that an increase in user activities (inner or incoming to RoIs) and the number of tweets can be used to infer geo-social events.

5.1.3 Semi-Supervised Approaches

A framework STED (semi-supervised system) is proposed to automatically detect and interactively visualize events of a targeted type (crimes, civil unrest, disease outbreaks) from Twitter [45]. Unlike most of the recent research work, this framework is mainly designed to target small-scale city-level or even street-level events by taking input as the topic of interest from users and retrieving tweets and detecting events of interest and summarizing the results in visual form. STED comprises of four steps: 1) automatic label creation and expansion: it collects domain-specific descriptions using the natural language toolkit (NLTK) and extracts named entities (noun) and action words (verb) from the media news description as a candidate query word set for tweets. Words of interest are given to the label generation module to retrieve tweets and mark them positive if each tweet contains at least one named entity and at least one action word. Social ties are then identified by building heterogeneous (i.e., term-tweet, hashtag-tweet) networks created subsequently one after another to filter out trivial and irrelevant terms; 2) customized text classifier for Twitter: the graph partitioning method is used to group the event-related terms that are left after the label generation module and mini-clusters are formed. These mini-clusters are further used in a specialized support vector machine (SVM) for final text classification. The important part of specialized SVM is feature selection. The authors introduced modified classic IRS features for their SVM by removing trivial and common terms, such as "people", "love" and most frequent terms to reduce the over-fitting issue; 3) enhanced location estimation algorithm: the social ties (using retweet RT, mentions @, and hashtags #) and spatial statistics, geo labels are produced. The authors hypothesized that tweets that do not contain geo-tags can be tagged if they contain similar social ties to those that have geo-locations; 4) a GUI is developed for visualization and analysis purposes that shows the major events along with their locations on a geographical map. The results are claimed to be significant and a lead time of 2.42 days ahead of traditional media.

Chen et al. propose a query-based event detection technique using user-defined keywords [28]. Initially, for keyword expansion, Baike corpus in the Chinese language is used to index the keywords. Other studies that focus on English content used Wikipedia [20] and web n-grams [84] for the same purpose. In relation to candidate expansion features, indexed Baike documents are retrieved along with term similarity score based on user-provided topic words. The top 200 terms appearing in the retrieved documents are considered candidate expansion terms. After keyword expansion, topic-related microblogs are filtered through expanded terms using a logistic regression classifier. The 2-class classifier separates positive and negative documents. Finally, heat word clusters are created to group terms based on their co-occurrence and frequency using the classic affinity propagation clustering algorithm. The final clusters represent events.

A system for Twitter-based Event Detection and Analysis (TEDAS) was developed by [66]. The three main operations of the system include: 1) detecting new events; 2) event ranking; and 3) temporal and spatial pattern generation for events. Tweets are collected using keyword-based rules related to crime and disaster events (CDE). A confidence measure is created to filter tweets which enhances the rule quality to improve CDE-related tweets and improve content coverage. Event detection is performed in two sequential steps that include offline and online processing. In offline processing, the crawled data is fed to the classifier to filter irrelevant tweets, then a meta information extractor extracts spatial and temporal information and stores this information along with the original tweet in an indexed database. In online processing, the system consistently retrieves emerging indexed data related to the queries provided by the user. Similar tweets based on their spatio-temporal pattern are grouped and ranked according to their importance.

5.2 Unspecified Event Detection (UED)

A few studies used supervised approaches for UED based on the nature of the content. However, most of the studies use semi-supervised or unsupervised approaches for UED. The following sections describe the details about detection approaches for UED.

5.2.1 Supervised Approaches

By exploring the characteristics of hashtags, Cui et al. reveal that hashtags follow power-law distribution [34]. Based on this behavior of hashtags, the authors raise three research questions: *"Do popular hashtags reveal breaking events?"*, *"Do popular hashtags indicate events or memes (conversational topics)?"* and *"Are popular hashtags contributed by the crowd?"* To answer these questions three attributes (i.e., hashtag instability, twitter meme possibility, and authorship entropy) were extracted. Each attribute is orthogonal and independent of each other. Considering L=Low and H=High value for each of the three attributes, the hashtag space is divided into eight sub-spaces by creating all possible combinations for all three attributes and each space is labeled with four possible classes as follows: A=Advertisements, M=Miscellaneous, T=Twitter Memes, and B=Breaking Events. The top 1% of the total hashtags are taken for the experiments and ground-truth is created by manually labeling the hashtags from two different annotators. A third annotator is involved when the first two annotators disagree. The label is selected based on the majority. In the absence of the majority, a hashtag is not included in the sample set. The subspace-based algorithm is then used to classify hashtags according to the four categories as mentioned earlier. The results are compared with popularity pattern algorithm [32] and compelling results are found.

5.2.2 Unsupervised Approaches

Most of the studies use unsupervised approaches for unspecified events by clustering contents into groups that have the characteristics of potential events. Zhang et al. [115] introduce a graph-based event detection technique using Twitter and Sina Weibo (a popular micro-blog in China). Data is collected from tweets generated by 313K randomly selected users from Twitter and 116K users participating in hot events recommended by the Sina Weibo micro-blogging service. The methodology for detecting events starts with the classic IR technique of tokenization and filtering stop-words. Micro-documents with less than four keywords are dropped. Each "word" is assigned a weight based on the TF, IDF, and user authority score. The Hidden Markov model is then used for the probabilistic automation (i.e. probability to produce related documents) of each word *low* and *high* where *low*=0 and *high*=1. Burst words with "high" significance are taken to generate a word relation graph. Nodes in the graph represent burst words and the edge/relationship weight is calculated according to their co-occurrence within each micro-blog, and the edge direction is detected by calculating the weight contribution from both nodes of the edge. The node that appears in most of the micro-blogs is considered the lead node. Edges with relatively small weights are removed. Strongly connected components in the graph are identified through graph clustering technique that uses the depth-first search algorithm and the connected components are considered as events. Sina Weibo recommends events which are labeled by site editors. These recommended events are taken as ground-truth for the data collection from Sina Weibo, whereas, in the case of Twitter no ground-truth is available for events. The events detected from Twitter data are manually observed by humans. The results are compared with two existing well-known techniques EDCoW and MSBI using MacroPrecision@K as an evaluation measure.

Zhou et al. develop a Bayesian model-based framework called the Latent Event and Category Model (LECM) [118]. Initially, a two-step filtering process is used. A lexicon is created via online news published around the same time as the tweets. Only those tweets that match any of the words from the lexicon are kept. Secondly, a binary classifier categorizes tweets into *event* and *non-event* classes using features like binary-words (word frequency ratio between event and non-event related tweets), news, time-phrase, opinionated words (manually selected phrases indicating events), currency, percentage sign, URLs, and mentions. Tweets are pre-processed based on temporal and linguistic features [40, 89]. After performing stemming, words with a frequency less than 3 are dropped. The named entities are then mapped on to semantic classes using freebase API. LECM assumes that each tweet is associated with an event (modeled as joint distribution of keywords, named entities, location, and time). LECM groups events into different event clusters. Each cluster is then analyzed to obtain a semantic class/label based on the cluster's entities and is employed as an event.

Zhou and Chen suggest that the problem of integrating ambiguous views from different users is not well-investigated [119]. They proposed a framework called Variable Dimensional Extendible Hash (VDEH) to detect composite social events over streams which fully utilizes the information of social data over multiple dimensions. By using a location-time constrained topic (LTT) model, the time, location, and tweet contents are captured. Then, the events are identified by conducting efficient similarity joins using a similarity query algorithm over social media streams. A series of experiments are conducted to prove the efficiency of LTT by comparing it with Online LDA-based event detection (OLDA).

To determine a useful method for detecting events that are spatial-temporal in nature, two different techniques are used to detect general retrospective events in Twitter stream [29]: 1) latent dirichlet allocation (LDA); and 2) space-time scan statistics (STSS). Initially, the content of tweets are used to classify them into topics using LDA, then the space-time permutation model (STPM) from STSS is used to create clusters with respect to space and time regardless of the content of the tweets. It creates a cylindrical window with a radius as space and height as the time over the geo-map. This process is repeated until all space time locations in the data have been visited. Each cylindrical window is viewed as a candidate for clusters based on their significance value. The topics discovered by LDA are then mapped on to spatial-temporal clusters.

Abdelhaq et al. in their study focus on localized events and create a system called *EvenTweet* [2]. Their technique detects localized events in real-time by continuously processing a live Twitter stream using a time-based sliding window approach. The event detection process is triggered every time a new sliding window appears. In a focused sliding window, the process identifies bursty keywords, then spatial distribution is estimated. The entropy for spatial signature is calculated to identify suitable bursty keywords for events. Low entropy of a keyword means the keyword has appeared in few locations and is thus important for detecting a localized event. Then, a single-pass clustering algorithm using cosine similarity is used to cluster selected candidate keywords based on their spatial signature. Clusters are then ranked based on their significance score and the top-K clusters are considered to represent events. The detected events are then highlighted on a geographical map with the help of a visual interface.

Gao et al. propose a detection method for geographical social events [38]. Initially geographical temporal patterns are created by counting the occurrences of tweets in specific regions by dividing the 24-hour day into 4 non-overlapping equal partitions of 6 hours each. Unusual geographical areas are discovered by normalizing the tweet count in each region and setting a threshold to mark a geographical area as unusual. Tweets from unusual geographical areas are then taken and a term vector is created using tf.idf. Adaptive k-mean clustering is applied with cosine similarity to create clusters. The clusters with a high frequency of tweets are considered as social events.

Liu et al. in their research study [67] investigate the detection of bursty events. Their proposed technique is focused on the post-processing of clusters to filter out noisy content and non-event clusters. As a pre-processing step, URLs and mentions are removed and retweets are merged with original tweet content. Unigram features are then extracted which include only nouns and verbs because of their importance in describing events. Then, a single-pass clustering algorithm is used to group bursty items. After creating the clusters, a filtering mechanism is performed to drop the clusters formed by advertisements, such as game promotions or product marketing. The filtration process involves quantifying the redundancy of cluster items in the current time window with respect to the previous one. If the redundancy ratio is decreased to a certain threshold, then the event cluster is seized, hence the final clusters are identified as bursty events.

Li et al. argue that for event detection, tweet segments (phrases), rather than unigrams (keywords) are more accurate since tweet segments have less noise and more meaningful units comparatively [62]. A framework is proposed with three steps: 1) tweet segmentation: an algorithm [63] to create tweet segments is used for tweets, which splits tweet content into segments. The algorithm provides an optimized solution for segmentation and splits the content unless the stickiness score (i.e., a statistical measure that shows the qualifying score for n-gram where $n \geq 2$) starts to reduce. 2) event segment detection: since the data is dynamic and continuously changing, the bursty segments along with the consideration of the users' frequency within a time window highlight the possibility of hot events. 3) event segment clustering: the content of tweets along with temporal patterns are used to measure the similarity among events. Event segments from previous steps are then grouped into clusters using their content-temporal similarity measure. The authors explain that many clustering algorithms can be applied, but they used Jarvis-Patrick algorithm [49]. They use a filtering mechanism by defining a statistical measure named *newsworthiness* to eliminate clusters that are not related to real-world events, and remaining clusters are identified as events. They also conclude that the newsworthiness to select relevant content is useful, and as well as features based on the tweet segments perform much better against unigrams for detecting events. A similar conclusion is presented by Aiello et al. in their study [4] regarding the usefulness of n-grams to detect social events from text streams such as Twitter.

5.2.3 Semi-Supervised Approaches

Huang et al. propose a framework called High Utility Pattern Clustering (HUPC) which combines k-nearest-neighbor and modularity-based clustering algorithms [46]. The framework first identifies high utility patterns from microblog streams using a pattern mining algorithm. Patterns are then sorted in decreasing order with respect to their *support* (based on association rule mining). Then, the top-K highly similar patterns are grouped using incremental clustering to represent emerging events. The incremental clustering of tweets has also been used to identify topics in ongoing discussions on Twitter [77], especially when coping with the scalability and performance issues.

Alsaedi et al. propose an online combined (classification-clustering) framework for disruptive event detection [10]. The framework adapts traditional IR techniques for pre-processing and creates a term vector using tf.idf. After pre-processing, the tweets are classified into the event and non-event categories using a Naïve Bayesian classifier to reduce noise and filter non-event tweets. A vast number of features, that include spatial (geo-tags, GPS locations, and inferred locations using named-entity recognition), temporal (tweet time, sliding window) and textual (near duplicates, retweet ratio, mention ratio, hashtag ratio, URLs, semantic category, semantic noun), are extracted for online incremental clustering. During the clustering process, the feature vector is calculated with 60-minute time window and 100 miles for location variance. The results show 84.18% precision. As an extension to their study, they evaluate the best suitable features for detecting disruptive events in [9] and find that it is not suitable to consider temporal, spatial or textual features in isolation to each other. The combination of these features leads to better event detection results.

A framework to detect emerging topics regarding specific organizations was developed in [27]. Data is gathered using three different sources *keyword source* (organization related keywords), *account source* (manually identified user accounts) and *org. key user source* (organization related users' published data). The single-pass incremental clustering algorithm [17] is used to handle a continuous stream of new tweets for topic detection. To infer the importance of the topic, two features *Topical user authority* and *Topical tweet influence* were formulated depending upon the users' tweet, users' retweet, followers and tweets' retweet counts, and tweets posted by authoritative users. To detect emerging topics, six key features (i.e., user increase rate, tweet increase rate, retweet increase rate, overlap in org. key users and influential users, overlap in keywords and influential topic keyword and increase rate in tweets' influential weights) are also extracted. In the absence of a benchmark, ground-truth was created by aligning the dataset with online news and manually labeling the collection of tweets. The detection model consisting of three voting-based ensemble classifiers Decision Tree, SVM, and Naive Bayesian, is chosen to learn from training data and detect emerging topics. The authors compared their results with three existing techniques TwitterMonitor(TM) [74], Topics over Time (TOT) [52] and NN-Dist [110] and found convincing performance of their co-learner ensemble classifier.

Tu and Ding developed a system that can automatically detect events and can process a large volume of tweets [103]. To reduce the data size by filtering irrelevant tweets, terms are ranked using tf.idf and top-K terms are taken as the feature vector. Bayes classifier is then used to categorize the tweets into multiple classes with an accuracy of 80%. After classification, the probability of each tweet is calculated against all other classes. Tweets which have no significant difference in probabilities for various classes are excluded with an assumption that tweets describing an event may relate to one class but not many. Then, the top-K categories clustering algorithm is used to group incoming tweets in the stream. The centroid of each cluster is calculated by taking the top-K tweets in each cluster which reduces processing overhead. The results show an improvement in accuracy and efficiency compared to the methods using single-pass clustering algorithm [2, 27, 67].

Supervised and semi-supervised techniques are resource consuming when categorizing unspecified events. TWICAL [90], an open-domain event extraction system, extracts multiple events based on the given event type, event phrase, date, and named entities. On a stream of tweets, the system tags tweets with part-of-speech using the named entity tagger technique described in [89] and extracts the named entities in association with event phrases. The model was trained by manually labeling 1000 tweets with event phrases and using a conditional random field (CRF) method to detect events and obtain F-score of 0.64. The extracted events are then associated with time using an algorithm proposed by [71] at an accuracy of 94% over 268 number of sampled events that were extracted earlier. Lastly, the latent variable model is adopted to identify event types. Events are then ranked using simple frequency and strength of association between the entity and calendar date to determine their significance using a statistical method Loglikelihood ratio (G^2).

5.3 Research Trends and Practices

Generally, there are two techniques: *Document Pivot* and *Feature Pivot* for detecting event and related topics. In document pivoted techniques, the documents are clustered by directly measuring their similarity with neighbors. However, feature pivoted techniques cluster important keywords representing event-related information. The issue with document pivoted techniques is that not all documents are related to events as it is assumed in the topic detection and tracking challenge. Therefore, the document pivoted techniques need arbitrary threshold for inclusion of a new document to the event clusters. On the other hand, feature pivoted techniques are based on measuring associations between terms and clustering, but often capture misleading term correlations due to the bias of burstiness [4, 81].

Detection techniques start with data filtration under certain criteria to remove unwanted noise and irrelevant content, then dissect tweet content for feature extraction and representation. Event detection techniques use one of the following approaches: supervised, unsupervised, or semi-supervised. Clustering using graph-partitioning [45, 112, 115] and state-of-the-art clustering [27, 61, 83, 94] are mainly used in unsupervised approaches. Twitter data is not labeled and is huge in size, thus clustering has an advantage because it can work on unlabeled data. Scalability is an issue in clustering-based approaches. To avoid the scalability issue, one fundamental solution is to reduce irrelevant and noisy data. Many studies [9, 10, 27, 115] adapt hybrid approaches and perform clustering in two steps to overcome this issue. First, a classifier is trained to categorize event and non-event tweets, then the remaining pre-processed data is used for clustering. Another way of reducing irrelevant data is through extracting a control vocabulary from the Wikipedia corpus [20, 59, 62], online published news [62, 118], or web n-gram [84]. Tweets below a similarity threshold are omitted, as it is hypothesized that the omitted data is not related to the event. Further details about filtration are discussed in Section 3.2. On the other hand, supervised approaches need labeled data. A classifier is trained over sample tweets set and is then applied on the complete data set [10, 54, 59]. Supervised approaches can be used with the assumption of a static environment by specifying a pre-defined set of events and classifying those events. Therefore, they are limited in scope. Secondly, Twitter data is dynamic, and new hashtags and keywords are emerging rapidly with the passage of time to report events. Adoptive incremental learning approaches should be developed to comprehend changes that may occur over time.

Semi-supervised approaches use a combination of unlabeled and labeled data. These approaches take a small set of labeled tweets to exploit the latent structure and feature patterns in the data. The unlabeled data is then processed to discover clusters as event-related information. These approaches can be used to avoid the limitations of supervised and unsupervised approaches [27, 45, 90]. Tables 8 and 7 summarize the details of event detection techniques used in various studies.

Event detection approaches (supervised, semi-supervised, and unsupervised) are not strictly bound to the types of events, however unsupervised approaches in this regard are very common. Obtaining benchmark or labeled data is complex and time consuming due to the huge volume of tweets [39], therefore most studies use unsupervised techniques for event detection. On the other hand, it is difficult to conclude that unsupervised approaches are superior. Most studies use unsupervised approaches for unspecified events by clustering content into groups that exhibit the characteristics of potential events. It is logical that when events and event-related information is unknown, then a better approach could be in the form of unsupervised or semi-supervised learning that mostly use cluster analysis and then identifying the clusters which relate to the events.

Based on the trends and common practices in this research area, we have developed a general *Event Detection on Twitter* (EDoT) framework as a guideline. This framework will help researchers to utilize Twitter data for event detection. It consists of four steps: 1) data acquisition; 2) feature extraction 3) event detection method; and 4) event representation. Each step is further elaborated in relation to the current research trends. The framework is shown in Fig. 4.

Based on the existing studies, data acquisition is discussed in Section 3, feature extraction in Section 4, and detection methods in Section 5. Event representation is a way to summarize and present event-related information to the end user interface. Each of the four steps involves multiple processes, and they are labeled with one or more of five tags. The tags show the apparent nature of each process. A description of each tag is given in Table 9. The processes with tag "+" are optional and may be dropped depending on the design decision of a research study. On the other hand, the processes with tag "#" are part of the implementation of the approach. The EDoT framework is self-descriptive, easy to understand, and will help in designing the architecture and models for event detection research.

6 Tools and Systems

In this section, we discuss the available systems and tools which have been developed as a result of

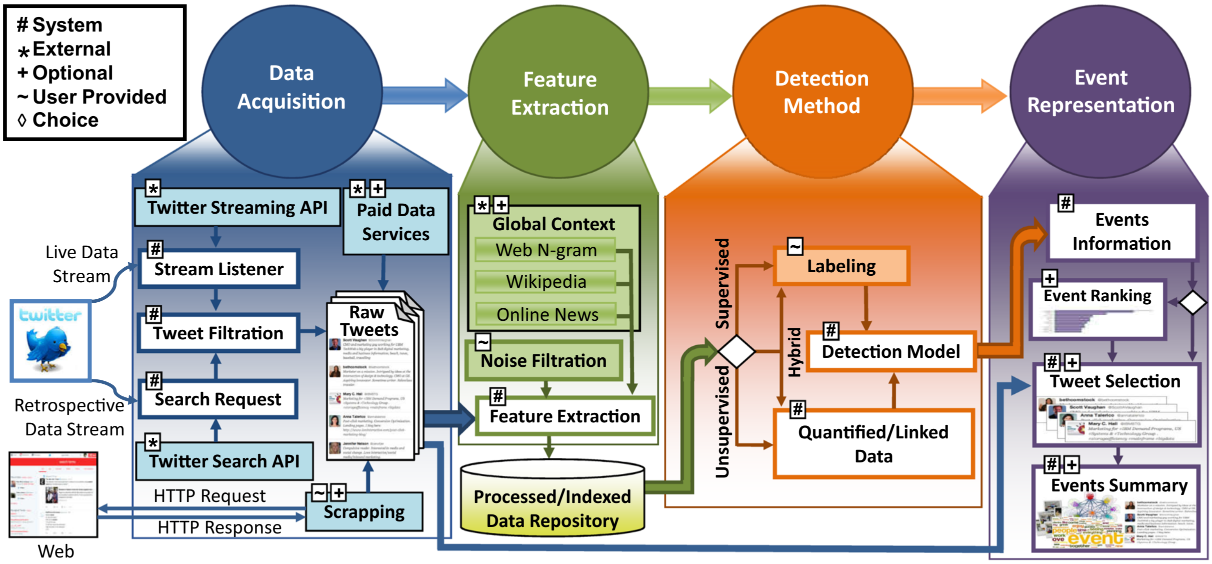

Fig. 4 A general Framework for event detection on Twitter

various studies. Some of the tools like *EnBlog* [11] and *TempEx* [90] were originally developed for online blogs and online news, respectively. Other researchers used them as building blocks to develop more complex tools to enhance their work that fits Twitter content. There are not many tools available publicly to analyze and detect events. Table 10 shows a list of tools developed and used in recent studies for event detection. TweetMotif [55] extracts a set of trending topics for the tweets retrieved against user query. The user interface provides a faceted search to aid searching tasks. The results are divided into two columns on the interface. The tool list down extracted topics related to the search query in the left column, while the right column shows actual tweets related to each extracted topic allowing a user to explore multiple topics at once. A similar system named Twitinfo [72] takes user query to generate trending topics in real-time by detecting peaks in the frequency signal of tweets. On the interface, the signal is visualized in real-time and peaks are given interactive labels. User can interact with detected peaks in real-time to explore event-related tweets. The interface also shows actual tweets with two colors representing the sentiment (i.e., red = negative and blue = positive). The TwitterStand system [94] is designed to detect news from the Twitter stream. The system uses two different sources of information that includes Gardenhose (Twitter API providing 10% tweets) against controlled search terms, and BirdDog API service for receiving targeted streams of two thousand users. It clusters the newsworthy tweets and further group together tweets according to their geo-location within the clusters. Another system named Twitter Intelligent

Table 9 Framework's process tags and their description

S.No	Tag	Name	Description
1	*	External	Available as an external resource in the form of API, software module or document corpus
2	#	System	System-based software module available as an integral part of the framework
3	+	Optional	Optional process, based upon the requirements it may or may not be used
4	~	User Provided	User provided software module, rules, data contents
5	◊	Choice	Choose any one of the outgoing paths

Table 10 Tools that are developed and used in different event detection studies

Reference	Event detection tool	Data Source	Remarks
Aiello et al. [4]	TMM	Twitter	Sensing trending topics in targeted events. Source available at (http://www.socialsensor.eu/results/software/87-topic-detection-framework)
Alvanaki et al. [11]	EnBlogue	Online Blogs	
Cheng and Wicks [29]	SatScan [57]	Twitter	Spatial, temporal and space-time statistical tool
Fujisaka et al. [37]	Geographical microblog collecting system	Twitter	
Hua et al. [45]	STED	Twitter	Targeted events
Krieger and Ahn [55]	TweetMotif	Twitter	Query based targeted events. Available at (http://tweetmotif.com/), Source (https://github.com/brendano/tweetmotif)
Lampos and Cristianini [59]	Flu Detector	Twitter	Available at (http://twitter.lampos.net/epidemics/)
Lee and Sumiya [61]	Geo-social monitoring system	Twitter	In continuation to S.No 3, the tool was developed for geo-social monitoring
Li et al. [62]	Twevent	Twitter	
Li et al. [63]	TwiNER	Twitter	Early crisis detection and response on target streams
Marcus et al. [72]	TwitInfo	Twitter	Query based targeted events with the user sentiment. Available at (http://twitinfo.csail.mit.edu/). Source (https://github.com/mitdbg/twitinfo)
Rill et al. [88]	PoliTwi	Twitter	Detects political emerging events. Available at (http://us.politwi.de/ and http://www.politwi.de/)
Ritter et al. [90]	TWICAL	Twitter	Detects event and infer the event date and time. Available at (http://statuscalendar.com. Alternative URL[a])
Ritter et al. [90]	TempEx [71]	Online news	Inputs a reference date, textual query, and parts of speech and marks temporal expressions with unambiguous calendar references
Ritter et al. [90]	Twitter-tuned POS tagger [89]	Twitter	Twitter-trained POS tagger
Sankaranarayanan et al. [94]	TwitterStand	Twitter	Newsworthy topic-detection from live stream Available at (http://twitterstand.umiacs.umd.edu/hyw/)
Valkanas et al. [105]	Twitter-ISA	Twitter	Traffic and flood-related event detection in real-time. Available at (http://www.insight-ict.eu/) for commercial use only.
Weng and Lee [112]	EDCoW	Twitter	Event detection with clustering of wavelet-based signals

[a]https://web.archive.org/web/20171112075526/http://ec2-54-170-89-29.eu-west-1.compute.amazonaws.com:8000/ (accessed on January 28, 2019)

Sensor Agent (Twitter-ISA) [105] detects traffic and flood-related incidents in real-time. The system gathers tweets using event-related control vocabulary. If a tweet is not geo-tagged, the system looks for location references in the tweet content and tags the location coordinates. Later, based on their geo-locations, a text classifier separates traffic and flood events. Twitter-ISA is developed under the INSIGHT[14] project and only available for commercial usage. Cheng and Wicks [29] developed a space-time permutation model to detect unspecified-unknown-general events using SaTScan[15] which was originally created by Kulldorff [57]. SaTScan is a free tool that examines temporal and spatial data. Similarly, for unknown-general events, *TWICAL*[16,17] was developed not only for event detection but also to associate events with

[14]http://www.insight-ict.eu/ (accessed on January 28, 2019)

[15]http://www.satscan.org/ (accessed on January 28, 2019)

[16]http://statuscalendar.com (Service is down, accessed on January 28, 2019)

[17]https://web.archive.org/web/20171112075526/http://ec2-54-170-89-29.eu-west-1.compute.amazonaws.com:8000/ (accessed on January 28, 2019)

calendar entries [90]. TWICAL also uses a statistical method G^2 to rank events according to their significance. Another tool, *Twitter-tuned POS tagger* [89], is used to extract and associate named entities with event phrases, furthermore, it finds unambiguous calendar references using TempEx [71]. To analyze emerging political trends, *PoliTwi*[18] was created [88]. PoliTwi collects data against keywords related to specific political events and visualizes emerging trends in different graphical forms. Similarly, another tool *Twevent* was developed for specified events on targeted Twitter streams [62]. The tool detects events based on tweet segments and by named entity recognition using another tool *TwiNER* [63]. A *Geo-social Monitoring System* [61] was developed to detect events based on crowd behavior in a geographical location. The system is an enhancement of the *Geographical Microblog Collecting System* [37] which discovers user movement patterns with geo-tagged content. The Geo-social Monitoring System was developed to detect events related to festivals and monitor crowd movements in different geographical regions of interest. Lampos et al. [59] in their study developed a *Flu Detector*.[19] Using Twitter content, the Flu Detector calculates the influenza-like illness (ILI) rate in different regions of the United Kingdom. A Twitter-based Event Detection and Analysis (TEDAS) system was developed in [66]. It works with both *online* and *off-line* tweet data. In offline processing, TEDAS continuously collects tweets with defined keywords for crime and disaster events, and extracts meta information to classify them in an indexed repository. Online processing involves user input with keywords and spatio-temporal information. The results are then visualized on graphical user interface.

A tool called TMM (Package,[20]Evaluation Script[21]) was developed and published as a software package [4]. TMM finds trending topics from targeted data streams. The basic purpose of the package is to support the research community in the design and development of efficient techniques for topic and event detection. TMM is also useful for the evaluation and comparison of different state-of-the-art approaches. The package has six well-known approaches implemented.

7 Discussion on Detection Techniques: Trends, Challenges and Future Directions

In this section, we discuss the shortcomings of the existing techniques and possible solutions to address these limitations. There are many studies on event detection on different social media platforms and in a variety of contexts. Due to the dynamic nature of Twitter content, the techniques which are used for traditional online media are not appropriate for Twitter. New techniques are introduced rapidly since Twitter has become a popular microblogging service. Hypothetically, it is assumed that all the documents used in the experiments are related to certain events. In the case of Twitter, this is not true because along with event-related information, the Twitter stream is flooded by irrelevant content. Almost 40% of the tweets, called babbles, are meaningless [82], and sometimes tweets spread rumors [24, 42] which introduce a significant amount of irrelevant content into the data.

Several studies label the data manually to address the challenge of irrelevant contents [54, 90]. Due to the huge size of the data, manual labeling is an expensive process. It is difficult to involve human experts directly to remove irrelevant tweets from a large-sized dataset [96]. A list of control vocabulary relevant to the targeted event(s) could be created, and then a binary-classifier could separate relevant and irrelevant tweets [118]. Another method is to label the data automatically by grouping the tweets using similarity measure and then labeling the cluster using bursty features from the cluster content [45]. However, the methods discussed above may not be used in the case of unsupervised approaches. Such approaches do not use class labels and highly rely on the filtration process to reduce the data using certain criteria, such as seed keywords [4, 51], word count [115], and term frequency [118].

Some of the challenges are due to the design of Twitter itself, that is, the limit of characters for a tweet compels users to convert their words into

[18] http://us.politwi.de/ and http://www.politwi.de/ (accessed on January 28, 2019)

[19] http://twitter.lampos.net/epidemics/ (accessed on January 28, 2019)

[20] http://www.socialsensor.eu/results/software/87-topic-detection-framework (accessed on January 28, 2019)

[21] http://www.socialsensor.eu/results/datasets/72-twitter-tdt-dataset (accessed on January 28, 2019)

abbreviations, and to use information and improper sentence structures which leads to spelling and grammatical mistakes. Different keywords may refer to the same event or vice-versa; therefore, traditional NLP techniques are not suitable for Twitter data, hence making it more challenging for text analysis [13, 118]. Unstructured text poses a great challenge in identifying event related information in the presence of noise content. Users can write anything, causing noise in the Twitter stream [90, 118] such as romanization, self-made abbreviations such as OMG (Oh my God), Bday (Birthday), ty (Thank you), using a diversity of hashtags, spelling mistakes, and multilingual tweets [30]. All these factors contribute to the degradation of the performance of detection approaches and make the event detection task challenging.

We observed a research gap concerning multilingual content. Chierichetti et al. proposed an approach which is language independent [30]. It focuses on non-textual features such as time and retweet. Other approaches, such as [29, 94, 112, 119] model the tweets as a bag-of-words and rely on burstiness. Such methods ignore the multilingual semantics. Hence, there is a need for robust techniques that can effectively work for multiple languages.

To address the problem of spelling mistakes, the classic IR technique for limiting words to their root form, or soft matching the terms using the Levenshtein similarity measure [4] can be used to overcome this challenge.

Processing an extensive amount of data from Twitter requires powerful computational machines [13]. Similar to event detection techniques, data preprocessing techniques also consume a significant amount of computational resources, hence making it difficult and challenging to design and develop systems that can be used easily by ordinary users on their PCs and smart devices to track events from a live stream. With the limitations of processing devices, the systems in [59, 88, 90] detect and track events in the context of a user-centered scenario, where the user focuses on targeted events by providing key-terms in the query, and the targeted stream is processed, hence reducing the amount of data and computational complexity.

Another challenging issue with Twitter data is benchmarking. The huge volume of data makes it a difficult task. A single benchmark coping with diverse events is very costly. Furthermore, the methods used for benchmarking are discussed in detail in Section 3.4.1.

Diverse real-life events are very dynamic with respect to their content. Diverse events differ in popularity, user participation, and content size (number of messages) [62]. Event detection techniques must cope with the diversity among different types of events on Twitter. Handling diversity and separating noisy data from real-world event information makes event detection a challenging task. Thus, efficient and scalable techniques are required to process such diverse raw data, especially in dealing with a real-time Twitter stream.

Data collection is time-consuming and costly. Free APIs provide limited access. Retrospective data cannot be acquired for more than the past seven-fifteen days. Similarly, the number of request to a live stream is also limited, which leads to a low data coverage issue. Data quality in terms of the fair coverage of event content is not guaranteed. On the other hand, full data coverage is not free and is not affordable for most researchers. We address data collection and related issues in detail in Section 3. In addition to the issue of data collection, datasets lack spatial information. Only approximately 2% of the total tweets are geo-tagged [45]. Working with geo-sensitive events is challenging in the absence of the required meta-data. Various studies [9, 10, 61, 90] address this challenge by extracting the geo-locations from the tweet content and tag the locations explicitly (see Section 4 for details).

Support for image and short video sharing in tweets increases the complexity of content when analyzing. Textual and multimedia content both have different paradigms of mining techniques, and to date, researchers have paid fewer attentions to multimode content analysis in Twitter. A few studies use text and images to detect events [8, 111]. One possible reason for this could be that support for images and videos has recently been integrated into Twitter, and much of the research work is focused on textual content and its social aspects. The possibility of contrary images and videos in tweets make it a great challenge to associate them with a single related event. However, the high-level features of images along with other tweet features might be helpful and may contribute to the improvement of the quality of the results.

In some studies like [118], redundant content due to retweets is left unhandled . Tweets generated by influential users (with greater followship) may affect the bursty features and add bias to the measure used for event detection or cluster labeling. These users' influence is useful when predicting an event's popularity [115], but can be misleading for event detection at the same time. The redundant content can be handled by devising a weighting criterion, as used in some of the studies [27, 30, 54, 66], to penalize the bias they may cause in the burstiness of event-related topics.

The techniques which are typically developed for small local events [61, 66], may not be useful for major outbreaks and global events. Comparative studies on the effectiveness of event detection techniques on different, small/local as well as major/global events are still a gray area.

Most of the techniques rely on the classic IR model and focus on the English language only. Non-English tweets, stop-words, and common words are normally filtered out or ignored. Data generated from various geo-locations have multi-linguistic contents expressing event-related information [115]. English is the language most used on Twitter, with 51% of tweets being in English. Japanese, Portuguese, Indonesian, and Spanish account for 39% of tweets [44]. A total 49% of tweets are in non-English languages that highlights the significance of techniques cope with multilingual contents. Event detection techniques that can handle multiple languages need more research focus to cover events on a broader scale. A method for translating multiple languages into a uniform language may help to increase the performance. Similarly, non-text based features [111] or the fusion of text and multimedia features [8] would also be useful to extract event-related contextual information.

8 Conclusions

Event detection on Twitter is an active research area. The availability of published content on Twitter makes it interesting for the monitoring, tracking, and detection of meaningful information which describe real-world events. The identification of escalated events is useful for many applications. In this survey, we covered event detection studies on Twitter along with related aspects and critically analyzed research parameters such as event detection models, techniques, case studies, feature extraction, evaluation, and the benchmarking of datasets. We provide an overview of the important issues and research challenges in this area and provide the EDoT framework as a guideline to researchers interested in this field. Efficiency and scalability are major concerns in this area. There are great opportunities to develop better feature extraction, filtering, and event detection techniques on microblogging services as well as combining different data sources of social and traditional media to improve the event detection techniques.

References

1. Abbasi, R.A., Maqbool, O., Mushtaq, M., Aljohani, N.R., Daud, A., Alowibdi, J.S., Shahzad, B.: Saving lives using social media: analysis of the role of Twitter for personal blood donation requests and dissemination. Telematics Inform. **35**(4), 892–912 (2018)
2. Abdelhaq, H., Sengstock, C., Gertz, M.: Eventweet: online localized event detection from Twitter. Proceedings of the VLDB Endowment **6**(12), 1326–1329 (2013)
3. Adedoyin-Olowe, M., Gaber, M.M., Dancausa, C.M., Stahl, F., Gomes, J.B.: A rule dynamics approach to event detection in Twitter with its application to sports and politics. Expert Syst. Appl. **55**, 351–360 (2016)
4. Aiello, L.M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Goker, A., Kompatsiaris, I., Alejandro J.: Sensing trending topics in Twitter. IEEE Trans. Multimedia **15**(6), 1268–1282 (2013)
5. Alhindi, W.A., Talha, M., Sulong, G.B.: The role of modern technology in arab spring. Archives Des Sciences **65**(8), 101–112 (2012)
6. Allan, J.: Introduction to topic detection and tracking. In: Topic detection and tracking, pp. 1–16. Springer (2002)
7. Allan, J., Carbonell, J.G., Doddington, G., Yamron, J., Yang, Y.: Topic detection and tracking pilot study: final report. In: Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, pp. 194–218 (1998). USA, Online
8. Alqhtani, S.M., Luo, S., Regan, B.: Fusing text and image for event detection in Twitter. CoRR arXiv:1503.03920, pp. 1–9 (2015)
9. Alsaedi N., Burnap, P., Rana, O.: Identifying disruptive events from social media to enhance situational awareness. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ASONAM'15, pp. 934–941. ACM, New York (2015)
10. Alsaedi, N., Burnap, P., Rana, O.F.: A combined classification-clustering framework for identifying disruptive events. In: Proceedings of the 2014 ASE SocialCom Conference. Stanford University, CA, USA, Online, pp. 1–10 (2014)
11. Alvanaki, F., Sebastian, M., Ramamritham, K., Weikum, G.: Enblogue: emergent topic detection in Web 2.0 streams. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, SIGMOD'11, pp. 1271–1274. ACM, New York (2011)

12. Asur, S., Huberman, B.A.: Predicting the future with social media. In: Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - vol. 01, WI-IAT'10. Washington, IEEE Computer Society, pp. 492–499 (2010)
13. Atefeh, F., Khreich, W.: A survey of techniques for event detection in Twitter. Comput. Intell. **31**(1), 132–164 (2015)
14. Becker, H.: Identification and characterization of events in social media. PhD thesis, Columbia University (2011)
15. Becker, H., Chen, F., Iter, D., Naaman, M., Gravano, L.: Automatic identification and presentation of Twitter content for planned events. In: International AAAI Conference on Web and Social Media. USA, AAAI, pp. 655–656 (2011)
16. Becker, H., Iter, D., Naaman, M., Gravano, L.: Identifying content for planned events across social media sites. In: Proceedings of the 5th ACM International Conference on Web Search and Data Mining, WSDM'12, pp. 533–542. ACM, New York (2012)
17. Becker, H., Naaman, M., Gravano, L.: Learning similarity metrics for event identification in social media. In: Proceedings of the 3rd ACM International Conference on Web Search and Data Mining, WSDM'10, pp. 291–300. ACM, New York (2010)
18. Becker, H., Naaman, M., Gravano, L.: Beyond trending topics: real-world event identification on Twitter. In: International AAAI Conference on Web and Social Media. USA, AAAI, pp. 438–441 (2011)
19. Benhardus, J., Kalita, J.: Streaming trend detection in Twitter. Int. J. Web Based Communities **9**(1), 122–139 (2013)
20. Benson, E., Haghighi, A., Barzilay, R.: Event discovery in social media feeds. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - vol. 1, HLT'11. Stroudsburg, Association for Computational Linguistics, pp. 389–398 (2011)
21. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: conversational aspects of retweeting on Twitter. In: 2010 43rd Hawaii International Conference on System Sciences. USA, IEEE, pp. 1–10 (2010)
22. Cai, H., Yang, Y., Li, X., Huang, Z.: What are popular exploring Twitter features for event detection, tracking and visualization. In: Proceedings of the 23Rd ACM International Conference on Multimedia, MM'15, pp. 89–98. ACM, New York (2015)
23. Casati, R., Varzi, A.: Events. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Stanford University, USA. winter 2015 (2015)
24. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on Twitter. In: Proceedings of the 20th International Conference on World Wide Web, WWW '11, pp. 675–684. ACM, New York (2011)
25. Chavoshi, N., Hamooni, H., Mueen, A.: On-demand bot detection and archival system. In: Proceedings of the 26th International Conference on World Wide Web Companion, pp. 183–187 (2017). International World Wide Web Conferences Steering Committee
26. Chen, C., Wang, Y., Zhang, J., Xiang, Y., Zhou, W., Min, G.: Statistical features-based real-time detection of drifted Twitter spam. IEEE Trans. Inf. Forensics Secur. **12**(4), 914–925 (2017)
27. Chen, Y., Amiri, H., Li, Z., Chua, T.-S.: Emerging Topic Detection for Organizations from Microblogs. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 43–52. ACM, New York (2013)
28. Chen, Y., Xu, B., Hao, H., Zhou, S., Cao, J.: User-defined Hot Topic Detection in Microblogging. In: Proceedings of the 5th International Conference on Internet Multimedia Computing and Service, pp. 183–186. ACM, New York (2013)
29. Cheng, T., Wicks, T.: Event detection using Twitter: a spatio-temporal approach. PloS one **9**(6), e97807 (2014)
30. Chierichetti, F., Kleinberg, J., Kumar, R., Mahdian, M., Pandey, S.: Event detection via communication pattern analysis. In: Proceedings of 8th International AAAI Conference on Weblogs and Social Media. USA, Association for the Advancement of Artificial Intelligence, AAAI, pp. 51–60 (2014)
31. Cordeiro, M.: Twitter event detection: combining wavelet analysis and topic inference summarization. In: Doctoral Symposium on Informatics Engineering, DSIE, vol. 56, pp. 1–16 (2012). Porto, Portugal, University of Porto
32. Crane, R., Sornette, D.: Robust dynamic classes revealed by measuring the response function of a social system. Proc. Natl. Acad. Sci. **105**(41), 15649–15653 (2008)
33. Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J.: # Earthquake: Twitter as a distributed sensor system. Trans. GIS **17**(1), 124–147 (2013)
34. Cui, A., Zhang, M., Liu, Y., Ma, S., Zhang, K.: Discover breaking events with popular hashtags in Twitter. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 1794–1798. ACM, New York (2012)
35. de Ridder, D., Tax, D.M.J., Lei, B., Xu, G., Feng, M., Zou, Y., van der Heijden, F.: Supervised learning. In: Classification, Parameter Estimation and State Estimation: an Engineering Approach Using MATLAB. 2nd, pp. 207–258 (2017)
36. Farajidavar, N., Kolozali, S., Barnaghi, P.: A deep multiview learning framework for city event extraction from Twitter data streams. arXiv:1705.09975 (2017)
37. Fujisaka, T., Lee, R., Sumiya, K.: Discovery of user behavior patterns from geo-tagged micro-blogs. In: Proceedings of the 4th International Conference on Uniquitous Information Management and Communication, p. 36. ACM, New York (2010)
38. Gao, X., Cao, J., He, Q., Li, J.: A novel method for geographical social event detection in social media. In: Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service, pp. 305–308. ACM, New York (2013)
39. Giachanou, A., Crestani F.: Like it or not: a survey of Twitter sentiment analysis methods. ACM Comput. Surv. **49**(2), 28:1–28:41 (2016)
40. Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan,

J., Smith, N.A.: Part-of-speech tagging for Twitter: annotation, features, and experiments. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers. Stroudsburg, PA, USA .Association for Computational Linguistics, vol. 2, pp. 42–47 (2011)
41. Gu, H., Xie, X., Lv, Q., Ruan, Y., Shang, L.: Etree: effective and efficient event modeling for real-time online social media networks. In: 2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), vol. 1, pp. 300–307. IEEE, Washington (2011). IEEE Computer Society
42. Gupta, A., Lamba, H., Kumaraguru, P.: $1.00 per RT #BostonMarathon #PrayForBoston: analyzing fake content on Twitter. In: 2013 APWG eCrime Researchers Summit. USA, IEEE, pp. 1–12 (2013)
43. Gupta, P., Jindal, R., Sharma, A.: Community trolling: an active learning approach for topic based community detection in big data. J. Grid Comput. **16**(4), 553–567 (2018)
44. Hong, L., Convertino, G., Chi, E.H.: Language matters in Twitter: a large scale study. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. USA, Association for the Advancement of Artificial Intelligence, AAAI, pp. 518–521 (2011)
45. Hua, T., Chen, F., Zhao, L., Lu, C.-T., Ramakrishnan, N.: Sted: semi-supervised targeted-interest event detectionin in Twitter. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1466–1469. ACM, New York (2013)
46. Huang, J., Peng, M., Wang, H.: Topic detection from large scale of microblog stream with high utility pattern clustering. In: Proceedings of the 8th Workshop on Ph. D. Workshop in Information and Knowledge Management, pp. 3–10. ACM, New York (2015)
47. Iyengar, A., Finin, T., Joshi, A.: Content-based prediction of temporal boundaries for events in Twitter. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom). USA, IEEE, pp. 186–191 (2011)
48. Izquierdo, E.: Social networked media: advances and trends. In: Proceedings of the 2011 ACM Workshop on Social and Behavioural Networked Media Access, pp. 1–2. ACM, New York (2011)
49. Jarvis, R.A., Patrick, E.A.: Clustering using a similarity measure based on shared near neighbors. IEEE Trans. Comput. **100**(11), 1025–1034 (1973)
50. Jarwar, M.A., Abbasi, R.A., Mushtaq, M., Maqbool, O., Aljohani, N.R., Daud, A., Alowibdi, J.S., Cano, J.R., García, S., Chong, I.: Communiments: a framework for detecting community based sentiments for events. Int. J. Semant. Web Inf. Syst. (IJSWIS) **13**(2), 87–108 (2017)
51. Kaleel, S.B., Abhari, A.: Cluster-discovery of Twitter messages for event detection and trending. J. Comput. Sci. **6**, 47–57 (2015)
52. Kasiviswanathan, S.P., Melville, P., Banerjee, A., Sindhwani, V. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 745–754. ACM, New York (2011)
53. Khan, H.U., Daud, A., Ishfaq, U., Amjad, T., Aljohani, N., Abbasi, R.A., Alowibdi, J.S.: Modelling to identify influential bloggers in the blogosphere a survey. Comput. Hum. Behav. **68**, 64–82 (2017)
54. Khurdiya, A., Dey, L., Mahajan, D., Verma, I.: Extraction and compilation of events and sub-events from Twitter. In: Proceedings of the the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology. Washington, DC, USA, IEEE Computer Society, vol. 01, pp. 504–508 (2012)
55. Krieger, M., Ahn, D.: Tweetmotif: exploratory search and topic summarization for Twitter. In: Proceedings of AAAI Conference on Weblogs and Social (2010)
56. Krishnamurthy, B., Gill, P., Arlitt, M.: A few chirps about Twitter. In: Proceedings of the First Workshop on Online Social Networks, pp. 19–24. ACM, New York (2008)
57. Kulldorff, M.: Satscan user guide for version 9.0 (2011)
58. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web, pp. 591–600. ACM, New York (2010)
59. Lampos, V., Cristianini, N.: Nowcasting events from the social web with statistical learning. ACM Trans Intell Syst Technol (TIST) **3**(4), 72 (2012)
60. Lanagan, J., Smeaton, A.F.: Using Twitter to detect and tag important events in sports media. In: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media. USA, AAAI, pp. 542–545 (2011)
61. Lee, R., Sumiya, K.: Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, pp. 1–10. ACM, New York (2010)
62. Li, C., Sun, A., Datta, A.: Twevent: segment-based event detection from tweets. In: Proceedings of the 21St ACM International Conference on Information and Knowledge Management, pp. 155–164. ACM, New York (2012)
63. Li, C., Weng, J., He, Q., Yao, Y., Datta, A., Sun, A., Lee, B.-S.: Twiner: named entity recognition in targeted Twitter stream. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 721–730. ACM, New York (2012)
64. Li, H., Wang, Y., Zhang, D., Zhang, M., Chang, E.Y.: Pfp: parallel fp-growth for query recommendation. In: Proceedings of the 2008 ACM Conference on Recommender Systems, pp. 107–114. ACM (2008)
65. Li, J., Tai, Z., Zhang, R., Yu, W., Liu, L.: Online bursty event detection from microblog. In: Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing UCC'14. Washington, DC, USA, IEEE Computer Society, pp. 865–870 (2014)
66. Li, R., Lei, K.H., Khadiwala, R., Chang, K.C.-C.: Tedas: a Twitter-based event detection and analysis system. In: 2012 IEEE 28th International Conference on Data Engineering (ICDE). USA, IEEE, pp. 1273–1276 (2012)
67. Liu, C., Xu, R., Gui, L.: Burst events detection on microblogging. In: 2013 International Conference on Machine Learning and Cybernetics, vol. 04, pp. 1921–1924. IEEE, Washington (2013)

68. Liu, X., Nourbakhsh, A., Li, Q., Fang, R., Shah, S.: Real-time rumor debunking on Twitter. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 1867–1870. ACM, New York (2015)
69. Liu, X., Wang, M., Huet, B.: Event analysis in social multimedia: a survey. Front. Comp. Sci. **10**(3), 433–446 (2016)
70. Long, R., Wang, H., Chen, Y., Jin, O., Yu, Y.: Towards effective event detection, tracking and summarization on microblog data. In: Wang, H., Li, S., Oyama, S., Hu, X., Qian, T. (eds.) 12th International Conference on Web-Age Information Management, WAIM 2011, Wuhan, China, September 14-16, 2011. Proceedings, pp. 652–663. Springer, Berlin (2011)
71. Mani, I., Wilson, G.: Robust temporal processing of news. In: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 69–76 (2000)
72. Marcus, A., Bernstein, M.S., Badar, O., Karger, D.R., Madden, S., Miller, R.C.: Twitinfo: aggregating and visualizing microblogs for event exploration. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 227–236. ACM (2011)
73. Massoudi, K., Tsagkias, M., de Rijke, M., Weerkamp, W.: Incorporating query expansion and quality indicators in searching microblog posts. In: Clough, P., Foley, C., Gurrin, C., Jones, G.J.F., Kraaij, W., Lee, H., Mudoch, V. (eds.) 33rd European Conference on IR Research on Advances in Information Retrieval, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings, pp. 362–367. Springer, Berlin (2011)
74. Mathioudakis, M., Koudas, N.: Twittermonitor: trend detection over the Twitter stream. In: Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, pp. 1155–1158. ACM, New York (2010)
75. McMinn, A.J., Moshfeghi, Y., Jose, J.M.: Building a large-scale corpus for evaluating event detection on Twitter. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, pp. 409–418. ACM (2013)
76. Metzler, D., Cai, C., Hovy, E.: Structured event retrieval over microblog archives. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 646–655 (2012)
77. Minhas, M.F.K., Abbasi, R.A., Aljohani, N.R., Albeshri, A.A., Mushtaq, M.: Intweems: a framework for incremental clustering of tweet streams, pp. 87:1–87:4. ACM, New York (2015)
78. Mitchell, A., Hitlin, P.: Twitter reaction to events often at odds with overall public opinion (2013)
79. Morstatter, F., Pfeffer, J., Liu, H., Carley, K.: Is the sample good enough? comparing data from Twitter's streaming api with Twitter's firehose. USA, AAAI. In: 7th International AAAI Conference on Weblogs and Social Media, pp. 400–408 (2013)
80. Nguyen, D.T., Jung, J.E.: Real-time event detection for online behavioral analysis of big social data. Futur. Gener. Comput. Syst. **66**, 137–145 (2017)
81. Panagiotou, N., Katakis, I., Gunopulos, D.: Detecting events in online social networks: definitions, trends and challenges. In: Solving Large Scale Learning Tasks. Challenges and Algorithms, pp. 42–84. Springer (2016)
82. Petrović, S., Osborne, M., Lavrenko, V.: 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. Stroudsburg, PA, USA. Association for Computational Linguistics, pp. 181–189 (2010)
83. Phuvipadawat, S., Murata, T.: Breaking news detection and tracking in Twitter. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). Washington, DC, USA, IEEE, IEEE Computer Society, vol. 3, pp. 120–123 (2010)
84. Popescu, A.-M., Pennacchiotti, M.: Detecting Controversial Events from Twitter. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 1873–1876. ACM, New York (2010)
85. Popescu, A.-M., Pennacchiotti, M., Paranjpe, D.: Extracting events and event descriptions from Twitter. In: Proceedings of the 20th International Conference Companion on World Wide Web, pp. 105–106. ACM, New York (2011)
86. Raza, A.A., Habib, A., Ashraf, J., Javed, M.: Semantic orientation based decision making framework for big data analysis of sporadic news events. Journal of Grid Computing, pp. 1–17. (2018). https://doi.org/10.1007/s10723-018-9466-y
87. Ribeiro, S.S. Jr., Davis, C.A. Jr., Oliveira, D.R.R., Meira, W. Jr., Gonçalves, T.S., Pappa, G.L.: Traffic observatory: a system to detect and locate traffic events and conditions using Twitter. In: Proceedings of the 5th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, pp. 5–11. ACM, New York (2012)
88. Rill, S., Reinel, D., Scheidt, J., Zicari, R.V.: Politwi: early detection of emerging political topics on Twitter and the impact on concept-level sentiment analysis. Knowl.-Based Syst. **69**, 24–33 (2014)
89. Ritter, A., Clark, S., Etzioni, O., et al.: Named entity recognition in Tweets: an experimental study. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 1524–1534 (2011)
90. Ritter, A., Etzioni, O., Clark, S., et al.: Open domain event extraction from Twitter. In: Proceedings of the 18Th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1104–1112. ACM, New York (2012)
91. Saeed, Z., Abbasi, R.A., Razzak, M.I., Xu, G.: Event detection in Twitter stream using weighted dynamic heartbeat graph approach. IEEE Computational Intelligence Magazine. (2019). In-Press. arXiv:1902.08522
92. Saeed, Z., Abbasi, R.A., Sadaf, A., Razzak, M.I., Xu, G.: Text stream to temporal network - a dynamic heartbeat graph to detect emerging events on Twitter. In: Proceedings of the 22nd Pacific-Asia Conference Advances in

Knowledge Discovery and Data Mining, PAKDD 2018, Melbourne, Australia, June 3-6, 2018 (2018)
93. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web, pp. 851–860. ACM, New York (2010)
94. Sankaranarayanan, J., Samet, H., Teitler, B.E., Lieberman, M.D., Sperling, J.: Twitterstand: news in Tweets. In: Proceedings of the 17th acm Sigspatial International Conference on Advances in Geographic Information Systems, pp. 42–51. ACM, New York (2009)
95. Sedhai, S., Sun, A.: Semi-supervised spam detection in Twitter stream. IEEE Transactions on Computational Social Systems **5**(1), 169–175 (2018)
96. Sethi, T.S., Kantardzic, M.: On the reliable detection of concept drift from streaming unlabeled data. Expert Syst. Appl. **82**, 77–99 (2017)
97. Shah, A., Halim, Z.: On efficient mining of frequent itemsets from big uncertain databases. Journal of Grid Computing, pp. 1–20. https://doi.org/10.1007/s10723-018-9456-0 (2018)
98. Da Silva, N.F.F., Coletta, L.F.S., Hruschka, E.R.: A survey and comparative study of tweet sentiment analysis via semi-supervised learning. ACM Comput. Surv. **49**(1), 15:1–15:26 (2016)
99. Somanchi, S., Neill, D.B.: Graph structure learning from unlabeled data for early outbreak detection. IEEE Intell. Syst. **32**(2), 80–84 (2017)
100. Takahashi, T., Abe, S., Igata, N.: Can Twitter be an alternative of real-world sensors? In: Jacko, J.A. (ed.) Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments: 14th International Conference, HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part III, pp. 240–249. Springer, Berlin (2011)
101. Tang, L., Liu, H.: Toward predicting collective behavior via social dimension extraction. IEEE Intell. Syst. **25**(4), 19–25 (2010)
102. Troncy, R., Malocha, B., Fialho, A.T.S.: Linking events with media. In: Proceedings of the 6th International Conference on Semantic Systems, vol. 42. ACM, New York (2010)
103. Tu, H., Ding, J.: An efficient clustering algorithm for microblogging hot topic detection. In: 2012 International Conference on Computer Science and Service System, pp. 738–741. IEEE, Washington (2012)
104. Tzelepis, C., Ma, Z., Mezaris, V., Ionescu, B., Kompatsiaris, I., Boato, G., Sebe, N., Yan, S.: Event-based media processing and analysis: a survey of the literature. Image Vis. Comput. **53**, 3–19 (2016). Event-based Media Processing and Analysis
105. Valkanas, G., Gunopulos, D., Boutsis, I., Kalogeraki, V.: An architecture for detecting events in real-time using massive heterogeneous data sources. In: Proceedings of the 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining: algorithms, Systems, Programming Models and Applications, pp. 103–109. ACM (2013)
106. Valverde-Albacete, F.J., Peláez-Moreno, C.: 100% Classification accuracy considered harmful: the normalized information transfer factor explains the accuracy paradox. PloS one **9**(1), e84217 (2014)
107. Velampalli, S., Eberle, W.: Novel graph based anomaly detection using background knowledge. In: Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference, pp. 538–543 (2017)
108. Vieweg, S., Hughes, A.L., Starbird, K., Palen, L.: Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1079–1088. ACM, New York (2010)
109. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)
110. Wang, X., McCallum, A.: Topics over time: a non-Markov continuous-time model of topical trends. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 424–433. ACM, New York (2006)
111. Wang, Y., Kankanhalli, M.S.: Tweeting cameras for event detection. In: Proceedings of the 24th International Conference on World Wide Web, WWW '15. Republic and Canton of Geneva, Switzerland, International World Wide Web Conferences Steering Committee, pp. 1231–1241 (2015)
112. Weng, J., Lee, B.-S.: Event detection in Twitter. ICWSM **11**, 401–408 (2011)
113. Williams, S.A., Terras, M.M., Warwick, C.: What do people study when they study Twitter? classifying Twitter related academic papers. J. Doc. **69**(3), 384–410 (2013)
114. Yang, Y., Pierce, T., Carbonell, J.: A study of retrospective and on-line event detection. In: Proceedings of the 21St Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 28–36. ACM, New York (1998)
115. Zhang, X., Chen, X., Chen, Y., Wang, S., Li, Z., Xia, J.: Event detection and popularity prediction in microblogging. Neurocomputing **149**, 1469–1480 (2015)
116. Zhao, J., Wang, X., Jin, P.: Feature selection for event discovery in social media a comparative study. Comput. Hum. Behav. **51**, 903–909 (2015)
117. Zheng, Y., Li, L., Zheng, F.: Social media support for knowledge management. In: 2010 International Conference on Management and Service Science, pp. 1–4. IEEE, Washington (2010)
118. Zhou, D., Chen, L., He, Y.: An unsupervised framework of exploring events on Twitter: filtering, extraction and categorization. In: Proceedings of 29th AAAI Conference on Artificial Intelligence. USA, AAAI, pp. 2468–2475 (2015)
119. Zhou, X., Chen, L.: Event detection over Twitter social media streams. The VLDB Journal—the International Journal on Very Large Data Bases **23**(3), 381–400 (2014)
120. Zhu, X., Davidson, I.: Knowledge discovery and data mining: challenges and realities. IGI global, Hershey, PA, USA (2007)