Review

Redefining Event Detection and Information Dissemination: Lessons from X (Twitter) Data Streams and Beyond

Harshit Srivastava and Ravi Sankar *

iCONS Lab, Department of Electrical Engineering, University of South Florida, Tampa, FL 33630, USA; harshit@besensable.com

* Correspondence: sankar@usf.edu

Abstract: X (formerly known as Twitter), Reddit, and other social media forums have dramatically changed the way society interacts with live events in this day and age. The huge amount of data generated by these platforms presents challenges, especially in terms of processing speed and the complexity of finding meaningful patterns and events. These data streams are generated in multiple formats, with constant updating, and are real-time in nature; thus, they require sophisticated algorithms capable of dynamic event detection in this dynamic environment. Event detection techniques have recently achieved substantial development, but most research carried out so far evaluates only single methods, not comparing the overall performance of these methods across multiple platforms and types of data. With that view, this paper represents a deep investigation of complex state-of-the-art event detection algorithms specifically customized for streams of data from X. We review various current techniques based on a thorough comparative performance test and point to problems inherently related to the detection of patterns in high-velocity streams with noise. We introduce some novelty to this research area, supported by appropriate robust experimental frameworks, to performed comparisons quantitatively and qualitatively. We provide insight into how those algorithms perform under varying conditions by defining a set of clear, measurable metrics. Our findings contribute new knowledge that will help inform future research into the improvement of event detection systems for dynamic data streams and enhance their capabilities for real-time and actionable insights. This paper will go a step further than the present knowledge of event detection and discuss how algorithms can be adapted and refined in view of the emerging demands imposed by data streams.

1. Introduction

Microblogging is a digital communication method that allows users to disseminate brief messages, URLs, and multimedia content instantaneously to a network of followers, transforming interpersonal connections and information sharing. These intended messages in the form of tweets were formerly limited to 140 characters on X, a prominent microblogging network. Approximately 500 million tweets are disseminated daily by 400 million monthly active users of X. As a result, X has become an indispensable channel to access the latest information. This high activity has also attracted growing amounts of research interest in the study of its usage for various domains, including disaster response [1] and epidemic tracking [2]. In this regard, many new methods have been developed to tap into its data streams to discover interesting events that make X one of the best tools to analyze

dynamic real-time information. They generally adopt the event definition put out by Topic Detection and Tracking (TDT) research, which describes an event as a real-world occurrence that transpires within a given geographic area and time frame [1]. The primary objective of these event detection methods is to meet the specific needs of X data, including the length of tweets and the presence of considerable spam, typographical errors, and informal phrases, among other factors. While the majority of proposals include qualitative data supporting the technique's benefits, few provide a quantitative evaluation or contrast the technique's outcomes with those of its competitors. The absence of quantitative evaluations and competitive comparisons casts doubt on the effectiveness and reliability of these event detection methods. In the absence of definitive evidence of superiority, measuring the real worth of this strategy proves challenging.

Building on previous efforts and prior surveys [3–6], the present research introduces scalable assessment approaches for the quantitative evaluation of X event detection systems. In light of the caveats mentioned by Bontcheva and Rout [5] and Atefeh and Khreich [6] our main objective is to suggest standardized measures that can be directly used for assessing the results of existing and future event detection methods. These techniques bridge qualitative insights with measurable outcomes and provide a basis for runtime and task-specific performance assessment.

Our contributions include the following:

1. An all-inclusive analysis of the state-of-the-art event detection algorithms for X data streams, incorporating findings from multipolar methods that have not been thoroughly investigated and bridged in prior surveys [2–55].
2. Following on from the discussions of Nurwidyantoro and Winarko [3] and Alvanaki et al. [7], some new performance metrics are being developed to assess task-based correctness with execution efficiency in real-time settings.
3. A review of the existing techniques, pointing out their strengths and weaknesses, and possible improvements, in light of the works by and Meladianos et al. [23] and Guille and Favre [24] as sources of inspiration.

This work also contributes to general event detection by refining existing methods and introducing new, challenging evaluation metrics. We conduct a comprehensive evaluation of our work, situating it within the framework of contemporary methodologies.

2. Background on Social Data Sources

In recent years, several studies on event detection and tracking techniques for X have been published. Consequently, several surveys have been created to document the current state of the art. Nurwidyantoro and Winarko's [3] study outlines strategies for detecting disasters, traffic, diseases, and news events. Madani's work [4] addresses disease identification, natural disasters, trends, and public sentiment assessment which reviews many fields of event detection and the distinct techniques to address each challenge. Conversely, the fundamental perspective of Bontcheva and Rout arises from a comprehensive effort to comprehend social media data [5]. The survey's subjects cover modeling user and network behavior, as well as intelligent semantic information access. The research referenced several event detection systems based on clustering models and signal processing techniques. They also examine sub-event detection methodologies, noting that event detection can occasionally be intricate and convoluted. Atefeh and Khreich [6] categorize event detection strategies based on approaches, tasks, types of events, and applications in their critical assessment. They investigate the criteria employed for evaluating such systems, thereby offering a framework for comprehending event detection on social media. The surveys and studies illustrate that, although numerous methods for detecting current events are available, the majority utilize ad hoc performance measurements on hand-labeled datasets, with

few employing standardized evaluation procedures, most of which still necessitate a significant manual effort. Moreover, a limited number of the suggested strategies were directly contrasted with alternatives. Despite the increasing volume of research on this subject, the absence of comprehensive and comparative studies hinders the evaluation of strengths and limits in the event detection systems. This paper introduces novel evaluation criteria and a framework for enhanced scalability, automation, and comparability to resolve these concerns. It enhances event detection research by adopting a comprehensive integrated methodology for algorithm evaluation instead of solely providing task-specific solutions.

Although event detection research has achieved considerable progress, many critical gaps still remain. Among the most severe challenges is how to handle the characteristics of X data streams, such as the conciseness of tweets, informal language, and noise coming from spam and irrelevant content. All these often lead to inconsistencies in the preprocessing and analysis steps, affecting the performance of an event detection system. The dynamic nature of X data streams requires algorithms which can be adapted on the fly, a characteristic often missing with traditional methods relying either on static datasets or some pre-defined heuristic.

The most significant of the limitations of previous survey are on the metrics and methodologies used for the evaluation. Most of the studies, as discussed by Atefeh and Khreich [6], lack standardization of the metrics used, mostly relying on ad hoc metrics or manually labeled datasets that do not scale well to large datasets and are not generalizable across a wide range of applications. This has led to fragmented insights into the performance of different event detection techniques, thus making comparisons and evaluations difficult. A lack of ground truth datasets with robust benchmarks is contributing to a deficiency of systematic validation regarding the actual performance of such systems. To address these challenges, this paper proposes a modular, system-based approach that integrates advanced methodologies, such as semantic grouping, TF-IDF, and word embeddings, with scalable assessment measures. This framework, while targeted at dynamic and noisy data streams, allows for the easy integration of techniques such as the log-likelihood ratio and clustering-based anomaly detection. This technique aims to bridge qualitative insights with quantitative validation, focusing both on task-specific and runtime performance indicators.

Furthermore, our framework allows comparative analyses because it provides well-defined standards, which allowed us to compare various techniques for event detection under the same conditions. The contribution of this work is the development of a general methodology oriented toward scalability, flexibility in real time, and variety within the domain, taking the best from previous contributions [3,6,24]. This rigorous methodology covers the existing gaps but also paves the way for future investigations into the area of event detection systems based on X data streams. The following sections will first describe and assess the methodologies provided in previous publications on event detection strategies. Second, we offer accessible cooperation strategies for evaluating event detection algorithms.

2.1. Survey of Event Detection Techniques

Event detection is challenging in the presence of randomness and automation. To identify events, it is essential to recognize anomalies with a high precision and recall. This review analyzes the application methods employed in the Social News on the Web challenge [8] regarding submitted results, as well as the metrics of precision and recall, readability, coherence/relevance, and diversity. The results show how important it is to have a balance between automation and human control to improve accuracy and usefulness in many areas. Consequently, utilizing the Social News on the Web challenge, we offer

a technical assessment of alternative event detection and collaboration methodologies documented in the literature. These techniques aim to improve the precision and efficacy of event detection algorithms, hence enhancing the performance of social news platforms. Various ways to solve the problem from the pros to the cons can be learned from this study, comparing new research with the existing literature. Comparisons of that nature show what has gone wrong and what techniques seem to work in an ever-changing environment with enough data to move the area of event detection forward. This research establishes a basis for future progress in social analysis.

Table 1 summarizes some related works in the area of event detection techniques for X and other social media across a wide range of applications. It provides an understanding of the different methodologies used for social data collection techniques and their respective performance metrics. The "result parameters" column of the table presents the set of evaluation metrics utilized by these works to explain the performance of an event detection system in a number of real-world scenarios. The primary finding of this survey is that event detection precision is the most important metric, as 51% of the chosen research depends on it. The emphasis on precision highlights that accuracy lies in the proper identification of events within a data stream. The significance and dependability of identified events are directly dependent upon the accuracy of the model. In addition to accuracy, numerous studies also use supplementary metrics such as the F1 score, average precision, and area under the receiver operating characteristic curve, which quantify the equilibrium of false and true positives, thus providing a more nuanced understanding of model performance. Only two studies by, Alvanaki et al. [7] and Parikh and Karlapalem [46], have proposed a performance metric to measure efficiency with respect to the processing time and resource utilization of the algorithms. It is an important criterion because real-time event detection requires the implementation of systems that must process high volumes of social media data with sufficient speed and efficiency. Besides these traditional performance measures, some unconventional metrics have also been proposed in the literature. Alvanaki et al. [7] used the notion of relative accuracy that measures event-detection accuracy relative to some benchmark or reference level. Li et al. [9,10] and Guille and Favre [24] have used the DER, a metric that measures the occurrence of identical events detected over time, which becomes important for the long-term assessment of an event detector's performance. Task-based performance measures are significant; however, assessing the run-time performance of a technique is essential for determining its overall effectiveness. Incorporating measures such as relative accuracy and duplicate event rate enables researchers to achieve a comprehensive understanding of a technique's performance.

Table 1. List of event detection techniques.

Applications and Key Datasets	Papers	Challenges	Result Parameters
Disaster Management Datasets: Twitter API, Disaster-related Tweets (e.g., Earthquakes, Tsunami), Social Media Datasets	Srivastava et al. [1]	Data Noise	Influence and Precision Score
	Sakaki et al. [2]	Integration of geospatial model	Precision and F-Score
	Li et al. [9] and Li et al. [10]	Language Ambiguity	Precision Score
	Abel et al. [11]	Data Sparsity	Average Decision Score
	Adam et al. [12]	Dynamic Event patterns	Average Decision Score
	Terpstra et al. [13]	Vast Data Stream Management	Filtering Data from 100 K Tweets
	Nurwidyantoro et al. [3]	Feature Sets	Survey of Techniques
	Madani et al. [4]	Overlapping Event Signals	Survey of Techniques
	Winarko et al. [14]	Limited labeled Data	
	Aggarwal et al. [39]	Labor Intensive Annotations	Manual Tagging for Precision Score
	Phillips et al. [40]	Data Variability	Average Decision Score

Table 1. *Cont.*

Applications and Key Datasets	Papers	Challenges	Result Parameters
Disease Spread Datasets: Twitter, Flu Trends, Disease Outbreak Data (e.g., Zika, Influenza),	Nurwidyantoro et al. [3]	Noisy Signals, Overlapping Events and Feature Sets	Survey of Techniques
	Madani et al. [4]		Survey of Techniques
	Culotta [15]		Search of Correlation in Data
	Bodnar et al. [16]	Sparse Ground Truth	Correlation
	Ritterman et al. [17]		Filtering of Data of 48 million Tweets
	Wakamiya et al. [18]	Subtle Cues	
	Asgari-Chenaghlu et al. [19]	Noisy Correlation	
	Achrekar et al. [20]	Variable Reporting Rates	Search of Correlation in Data
	Alvanaki et al. [7]	Evolving Data Streams	
	Bontcheva et al. [5]	Too Complex Manipulations	Survey of Techniques
	Atefeh et al. [6]	Inconsistent Metrics	Survey on Evaluation Metrics
	Papadopoulos et al. [8]		Event Detection by Correlation
	Sankaranarayanan et al. [21]	Real-time constraint	Crawling and Spread metrics
Information Spread Datasets: Twitter, Reddit, Facebook, Wikipedia, Domain-specific Data	Walther et al. [22]	Overfitting	False Positive Detection
	Meladianos et al. [23]	Unbalanced Data	False Positive Accuracy
	Guille et al. [24]	Manual Cost	Precision and F-Score with Manual Tagging
	Petrović et al. [26]	Scaling up	Average Precision Score with Manual Tagging
	Marcus et al. [27]	Domain Specific Jargons	Precision Score
	Popescu et al. [28]	Dynamic Markets	Precision and F-Score
	Ishikawa et al. [29]	Filtering Spam	Crawling and Spread metrics
	Nishida et al. [30]	Large Scale Noise	Filtering of Data of 300 K Tweets
	Aiello et al. [31]	Subtle Patterns	Precision and F-Score
	Petrović et al. [32]	Manual Effort	Manual Tagging to obtain Precision Score
	Osborne et al. [33]	Dynamic Patterns	Time Taken for Information Spread
Business Analytics Datasets: Twitter, Amazon Reviews, Consumer Feedback, Marketing Campaign Data,	Benhardus et al. [34]	Evolving Slang	Precision and F-Score
	Cataldi et al. [35]	Data Variations Issue	Filtering of Data
	Lee et al. [36]	Annotation Bottleneck	Average Precision Score
	Mathioudakis et al. [37]	Trend Shifts	Crawling and Spread metrics
	Allan J. [38]	Broad Applicability	Filtering, Crawling, and Correlation
	Aggarwal et al. [39]	Labor Costs	Precision Score
Other Datasets: Multi-domain Data, Twitter, Public Web Data, Customer Support Data, Survey Data,	Cordeiro et al. [41]	Hetro Noises	Filtering and Reduction of Noise
	Ritter et al. [43]	Scaling Annotations	Precision Score
	Bahir et al. [44]	Ambiguity in Signals	Filtering of Data
	Martin et al. [45]	Balancing and Tuning	Recall Metrics of Activities
	Parikh et al. [46]	Resource Intensive	Filtering of Data
	Abdelhaq et al. [47]	High Noise	Filtering of Data
	Weiler et al. [48]	Method Selection	Survey of Techniques
	Corney et al. [49]	Complexity	Survey of Techniques
	Ifrim et al. [50]	Evolving Topics	Filtering of Data
	Zhou et al. [51]	Manual Overhead	Filtering of Data
	Thapen et al. [52]	Slow Interaction	Filtering of Data
	Monmousseau et al. [53]	Residual Noise	Filtering and Reduction of Noise
	Blei et al. [54]	Interpretablity	Concepts of Detection
	Hoffman et al. [55]	Scaling Models	
	McCreadie et al. [56]	Fast Changing Data	
	Cilibrasi et al. [57]	Complexity in Modeling	
	Khatoon et al. [58]	Hetro pattern	
	Bellatreche et al. [59]	Residuality	
	Savic et al. [60]	Ambiguity in Context	
	Jones [61]	Methodologies Evolution	

These techniques directly depend on the data collection method based on a set of two specific rules:

1. Users' direct following: This data collection process selects a default set of users which directly follow the users' streams of data irrespective of their locations;
2. Trends by global or geographical region: During this filtering process, the identification of trends is defined based on specific current topics in various applications with respect to geographic location or global level;

The rules mentioned above clearly state that the pool of social media data from which X and the other social media networks' data were compiled. The significant amount of

social media data available influences the evaluation methodologies employed in the research presented in Table 1, with numbers of tweets varying from 100,000 to 100 million. For instance, in [1], 1 million tweets were extracted via X's API, utilizing predetermined tags and keywords from January 2017 to January 2018, while Ritterman et al. [17] and Sankarnarayanan et al. [21] concentrated on domain-specific data and news data over two-month intervals, employing the user stream API of X. Furthermore, Smith et al. [25] executed a study on the X social media platform, aggregating 50 million tweets through a combination of topic searches and geolocation filters. This extensive dataset allowed for an in-depth analysis of user behavior and interactions on the site. The diverse quantities of social data collected in these studies underscore the necessity of accounting for data volume while formulating research procedures and deriving conclusions from social media data. Nonetheless, it is crucial to recognize that larger datasets do not invariably ensure a greater accuracy in outcomes.

The previously referenced survey has been organized according to the applications listed in Table 1, starting with a review of applications in disease spread tracking, disaster management, information dissemination, and business analytics, with the remaining applications categorized as others. The event detection approaches discussed in the reviewed papers were assessed according to the metrics specified in the results' parameters column. We are not assessing the accuracies of the current surveys, given the research presented demonstrates a strong sensitivity in detecting event occurrences. The findings are restricted regarding evaluators, and the majority of the work consists of manually categorized instances of occurrences. The absence of accuracy evaluations in the existing surveys may undermine the dependability of the proposed event detection algorithms.

Cullota [15] and Bodnar et al. [16] validated the influenza disease detection model through regression techniques to estimate disease spread while ensuring a high precision for the assessed datasets. In [2], 500,000 tweets were collected via a keyword-specific API during a duration of 28 months. The information collected was evaluated against multiple proposed models, achieving an accuracy of 0.78. On the contrary, Ritterman et al. [17] investigated the hypothesis that social media offers a framework for employing stock market forecasts to predict the spread of swine flu. The conventional regression classification method was employed for evaluation; however, the model was unable to identify the noisy data, resulting in inaccurate event predictions. In [18], a more objective methodology was employed to determine the propagation of influenza by integrating location aggregation and social media data, yielding a high accuracy of spread detection; however, the model exhibited a substantial dependence on location estimation variables. Asghari-Chenaghlu et al. [19] proposed the use of the transformer encoder in COVID-19 verification with data from social media, developing the concept of a universal word by applying some clustering techniques that indicate COVID-19 transmission for the small number of datasets developed from March to April 2020. In this work, their model is promising and provides better results in comparison to the usual method of detecting COVID-19. In addition, the transformer encoder methodology created a better and more efficient way of monitoring social media data to forecast virus spread. The research pinpoints that applying novel technologies and methodologies will be potentially useful for further developments in the earlier identification and tracking of infectious disease spread. At the same time, we have to keep in mind that the small dataset used may not be representative of the population in general, and hence can give biased results. When applied on a larger and diverse dataset, the accuracy may vary.

It becomes important to note here that keyword-specific and domain-specific event detection methods and algorithms are normally evaluated by comparing real-time data statistics. For example, the COVID statistics from John Hopkins may be used as the baseline

for evaluating how the disease has spread across different geographical locations or zip codes. Achekear et al. [20] clustered their survey data into 1000 clusters and compared it to the manually labeled data. This categorization helped to identify false positives and negatives, at the rate of 68% and 32%. Event detection has historically always contained clusters in the results due to its high sensitivity, which was observed during the detection of disease spread events by correlating the results with baseline survey data. In [7], a model was introduced to analyze the event detection of geo-location-specific tweets with a total of 22 million tweets collected through data crawling. The model possessed high accuracy in detecting anomalies with an accuracy of 0.89 compared to the manually labeled data, but it was incapable of detecting the different types of anomalies; here, it failed to pinpoint the events correctly with the accuracy dropping below 0.1. This problem was solved in [1] for the detection of different types of anomalies during hurricane Irma. Srivastava and Sankar [1] pointed out crucial steps in crawling data, detecting events, and labeling them as influences; then, these influences were further processed to detect the types of events with a high precision of 0.7. Their approach was based on a combination of machine learning algorithms and natural language processing techniques that could accurately classify the anomalies and spot the events correctly. With this focus on identifying influences in the data, the precision of the model in event detection improved manifold. By and large, their approach performed well in overcoming the drawbacks of the previous models and yielding a superior knowledge of events that occurred in the Twitter data.

Aggarwal et al. [39] presented an innovation in the field of disaster identification by segmenting the evaluation into two separate stages, providing a fresh perspective on the subject. The authors initially presented a case study to assess an unsupervised event detection model, highlighting the challenges posed by the streaming and unstructured characteristics of the data. The second part involved assessing a supervised model with self-generated ground truth. Precision and recall were calculated for two significant occurrences during this phase: the Japan Nuclear Crisis and the Uganda Protest. In the model pertaining to the Japan Nuclear Crisis, precision was 0.525 and recall was 0.62; during the Uganda Protest, precision was 1.0 while recall decreased to approximately 0.6. Also, the research was conducted on data that had already been prefiltered to a large extent, which raises some questions about their generalizability in real-world situations. Their method is quite suitable for model evaluation; however, it does not really reflect the real complexity of event detection from raw, unprocessed data originating from various sources. It was Phillips et al. [40] who introduced a remarkable approach to weather prediction concerning tornado events using X data, finding a strong correlation between detecting tornadoes by sentiment analysis alone and cross-referencing it with physical data. This result further underlines the use of unconventional sources of data in the timely identification of natural disasters. Their work, therefore, underlines the need for the integration of diverse data sources and sophisticated analytics to make predictions more accurate. Such an integration of more traditional meteorological data with sentiment analysis helped the researchers achieve a more holistic and accurate tornado forecast, hence proving interdisciplinary approaches to be effective for solving complex problems.

Notably, in their research on social sensors, the authors of [24] used the disaster detection method proposed by Sakaki et al. [2] for enhancing their system's detection capabilities. The system developed was particularly aimed at real-time event detection; for example, around 600 samples were used as the base sample of seismic events to train the model and classification methods were used to detect earthquakes with an accuracy of 0.66. The approach towards real-time event detection and identification is in line with the work of Becker et al. [62], who undertook extensive training over several weeks to evaluate the accuracy of the detected events, and then compared the results with those obtained from

manual evaluators in clusters. This evaluation resulted in a high detection accuracy while being insensitive to the structure of the underlying data. Their study set a standard where the system identified a narrative as an event. The model showed a high labeling capacity over a dataset of tweets, about 163.5 million gathered over six months, and proved capable of minimizing false positives as well as false negatives during an experiment in a streaming API. While the approach which Becker et al. followed might have worked fairly well for event detection, in actuality, their research lacks scalability and generalization testing using other datasets and cases.

The spread of news as well as the application of business analytics tools include techniques for identifying the popularity of different content and topics. The popularity of the concept is quantified on a real-time and an hourly basis using a trend crawling technique. In [24], the indexed topics considered popular were extracted from X using an algorithm to identify the probable popularity of a topic in the United States. The results of the event detection method were analyzed regarding X trends. These trends may be monitored continuously using the X API [10]. The method was evaluated in two stages. In the first stage of evaluation, there were no human interactions or manually annotated data involved, which resulted in a very low average precision value of 0.25. In the second part, manually labeled data were added and the average precision score increased to 0.65. However, papers [9,10] presented their evaluation on event detection with a clustering of wavelet-based signals (EDCoW). The EDCoW method builds signals for individual words by applying wavelet analysis to the frequency-based raw signals of the words. It then filters away the trivial words by looking at their corresponding signal autocorrelations. The results showed that the EDCoW method achieved an average precision score of 0.80, outperforming previous approaches by a significant margin. This demonstrates the effectiveness of using wavelet-based signals in the application of event detection and the importance of filtering out irrelevant words through signal autocorrelations. Overall, combining wavelet analysis and signal filtering in EDCoW turned out to be a promising approach to improving the accuracy of event detection tasks.

Li et al. [9] were the first to apply the EDCoW approach to event detection and an evaluation of their method showed a quite high sensitivity when assessed on an aggressively reduced dataset. This dataset consisted of tweets from the top one thousand users with large followings in Singapore and included tweets dated from 2010 onwards. This dataset was filtered down to a collection of 8140 unique phrases. Using this filtered dataset, EDCoW method was able to achieve an F-score of 0.76, again proving the goodness of fit of the event-detection technique in this particular filtered domain. The evaluation was mostly qualitative, focusing on a subjective analysis of the detected events rather than an in-depth quantitative comparison, thus limiting the interpretation of the method's applicability in general. Li et al. [10] present a comparison of the results of the EDCoW method against the results obtained using a segment-based event detection approach, using the same dataset as in [24]. The findings, however, showed that the segment-based approach gave a score of 0.86 for precision and thereby improved on the EDCoW methodology. The major shortfall of the segment-based approach, despite recording an improvement in the precision or memory scores, was that it achieved a memory score as low as 25% against the recall score, as reported in reference [24]. This discrepancy indicated that while the segment-based approach performed exceptionally well in identifying exact events, it performed poorly in incorporating more event types, especially those with limited representation in the training data. Li et al. [10] also realized that their segment-based approach has its limitations; in particular, on some specific event types that were poorly represented in the training dataset, it performed poorly. To this end, they suggested that increasing the size of the training corpus would provide a greater diversity of events, which in turn would enhance the

model's generalization ability for infrequent or niche events. They further suggested that using more advanced machine learning techniques would enhance both precision and recall and make the system more robust and accurate for event detection. Finally, they concluded by highlighting the importance of the continuous reevaluation and improvement of the event detection methods, underlining that, while the precision can be improved, reaching a high and constant value of recall, particularly for rare or unprecedented events, is one of the most important interests for future research. They encouraged using a fine-tuned methodology for training datasets and the introduction of sophisticated algorithms to enhance the reliability and the performance of the event detection systems for real and dynamic scenarios. The current review thus generates insights into the development of methodologies for event detection while also reflecting the need and importance of iterations of enhancements toward the realization of appropriate, reliable, and scalable event detection systems.

TwitInfo introduced in [27] is a technology designed to aggregate and visualize microblog data for the purpose of analyzing occurrences inside a social media stream. The developers of this technology employed manually labeled events, specifically for soccer matches, in their assessment and are expanding their efforts to include the detection of geological phenomena for catastrophe investigation. In the classification of soccer game events, TwitInfo's classifier scored 0.77 for precision and recall by correctly identifying 17 out of the 22 events, which is good accuracy in the identification of sports events. Event-related signals were identified with a precision of 0.14 in the case of major natural disasters, where six out of forty-four occurrences were detected, with a recall value of 1.0 because five of the events were correctly identified. The results showed that the peak identification algorithm in TwitInfo detected 80% to 100% of the peaks, which were manually labeled. This result underlines that a high precision is hard to attain for the identification of complex and low-frequency phenomena, such as significant disasters, which may be handled by more specialized algorithms handling diverse data features. Popescu et al. [28] have proposed a method for extracting events and their descriptions from microblogs. The study concentrated on distinguishing between events and non-events. A manually classified gold standard of 5040 images was utilized, categorized into events (2249) and non-events (2791). The outcomes of their methodology, Event Basic, included a precision of 0.691, a recall of 0.632, an F1 score of 0.66, an average precision of 0.751, and an average region of convergence of 0.791. Although these scores are commendable, the enhanced iteration of Event Basic, termed Event Aboutness, did not demonstrate substantial advancements; its results were comparable to the prior ones. This implies that their enhanced method failed to address certain basic shortcomings intrinsic to their event extraction strategy, although their strategy's performance was enhanced in the identification and classification of manually labeled peaks. Their findings highlight several constraints on the efficacy of their approach, particularly in attaining uniform performance enhancements across various kinds of events. These examples illustrate the difficulties and constraints of existing methodologies in event detection, particularly on social media and microblogging platforms. Although progress has been made in detecting specific types of events, considerable efforts are required to optimize these systems for diverse scenarios, intricate occurrences associated with natural disasters, or unstructured information inside dynamic data streams. Future research should concentrate on creating algorithms capable of adapting to these complexities to enhance precision and recall across various contexts, while also addressing scalability challenges associated with event extraction from extensive social media feeds.

Alvanaki et al. [7] carried out an evaluation that relied on public sentiment to decide if the results they reported were events. A website was created by the authors so that people could compare the results of their ENB method to a benchmark TM method [37].

Precision and performance at run-time were considered. ENB was significantly ahead in precision, reporting on average 2.5 out of 20 events, whereas TM reported only 0.8. The authors further considered the run-time efficiency, analyzing the influence of input size, algorithmic complexity, and hardware specification on the results. Consequently, it turned out to be difficult to make any meaningful conclusions based on the measurements. In their controlled experiments, they separated those variables to find the one that really affected how well the methods worked. By looking into how these parts work together, they provided useful information about how to make event detection systems work better and be more accurate. These studies laid the groundwork for later ones, which used more thorough, multidimensional evaluations to help make event recognition from dynamic data streams more scalable and reliable.

Aiello et al. [31] evaluated six topic detection techniques using three datasets of significant events from X (Twitter): BNGram, LDA, FPM, SFPM, Graph-based, and Doc-p. An exhaustive investigation was made possible by the datasets' variable temporal scales and subject turnover rates. They provided a thorough understanding of performance by presenting three primary evaluation metrics: subject recall, keyword precision, and keyword recall. BNGram has consistently provided the best subject recall among the approaches evaluated, making it resilient to the removal of numerous types of events. While some methods, like LDA, performed much better in capturing high-density events, they struggle in "noisy" conditions where overlapping or extra data are involved. For example, for the Super Bowl in the United States, which is a sporting event, standard topic detection techniques are likely to have a hard time accurately identifying relevant topics due to the high volume of noise from social media posts about commercials and halftime performances. This may spur researchers to develop more specialized algorithms that would be able to filter out irrelevant information and focus strictly on the main event itself.

Osborne et al. [33] conducted one of the first studies to look at latency in event recognition across platforms. They showed that events break first on X and then show up on Wikipedia after some time. Clearly, X has been found to be better than other traditional tools for reporting events in real time. Ritter et al. [43] suggested an open-domain event extraction system for X that was much more accurate than standard methods. Their method not only made event extraction more accurate, but it also let them see what events would happen in the future on a calendar. This showed how useful the system could be for finding events in real time and making predictions. In this way, you could obtain accurate and up-to-date event data, so it can be used in dynamic, real-time settings. Both studies show that X-based event recognition is becoming smarter. Osborne et al. looked at how timely it is compared to other sources, while Ritter et al. looked at both event extraction and prediction.

Martin et al. [45] performed a follow-up study to enhance the BNgram approach by determining its optimal window size and the most efficient combinations of grouping and topic-ranking methods. While their work did not directly enhance comparative assessments of event detection approaches, it provided significant insights. They discovered considerable heterogeneity in outcomes across several data sources like football and politics datasets. The memory rate for football-related issues exceeded 90%, significantly above the 60–80% range observed for politics, thus emphasizing the influence of event types and contextual variations on critical results. Their findings highlight the necessity of tailoring evaluations to certain event categories and data attributes, thus magnifying the heterogeneity introduced by the nature of occurrences. Whereas the ET system that Parikh and Karlapalem [46] constructed was assessed using two datasets, the VAST Challenge 2011 dataset, the VAST Challenge is a benchmark dataset designed for testing analytical and visualization tools. It typically includes complex, multi-faceted data scenarios that mimic

real-world problems. They used criteria for precision and recall that were consistent with one another and reported a precision of 0.91 and a recall of 21 for the VAST dataset. They were able to identify 23 events, two of which were considered to be meaningless. The efficiency of ET on smaller datasets that are well defined is demonstrated by the fact that it attained a precision of 0.93 and a recall of 14 out of 15 events that were discovered for the United States dataset. In addition, ET was able to handle 1,023,077 tweets in 157 s, which is equivalent to a throughput of 6516 tweets per second, which demonstrates the efficiency with which it processes information. This evidently shows that balancing the context in event detection is important to create robust and scalable techniques for real world situations and to remove bias.

A methodology that is anomaly-centric is utilized by MABED (Mention-Anomaly-Based Event Detection), which was created by Guille and Favre [24]. This methodology is utilized to identify events within noisy social media data. All of the evaluations showed that the method was superior to both ET and TS [53], displaying better robustness and accuracy, especially when pre-filtering the data for symbols like "@." The precision with which MABED is able to recognize subtle changes in the content of social media networks makes it particularly suitable for platforms that are prone to uncertainty and confusion, such as X (which was formerly known as Twitter). These findings provide more evidence that it is superior to traditional methods, which will make it easier to develop future event detection algorithms that are tailored to complex social media data. However, if we dig deeper, we find a study by Meladianos et al. [23], who created an event detection method that could detect sub-events with high precision and present detailed summaries of match occurrences. This was implemented on datasets consisting of soccer games. The study showed that their technique clearly outperforms the prior techniques on the sub-event detection task and highlights the efficacy of event-specific evaluation strategies. In addition, their algorithm managed to detect the majority of key sub-events during each match.

The transportation perspective on disease spread was presented by Monmousseau et al. [53]. They applied a fresh viewpoint to event detection by zeroing in on the transportation industry during the pivotal months of 2020 (February–March) when COVID-19 had affected the transportation industry and the illness was spreading. While most research focused on accuracy or memory measures for event detection in general, this one took a more nuanced approach by looking at how passenger-centric indicators like mood and empathy affected transportation choices made during public health emergencies. The accuracy of the detection of empathy is not explicitly measured, but this research does point to an understanding of these emotional factors as being the key to managing transportation systems under high uncertainty. However, since empathy and mood are relatively subjective elements in people, it is hard to generalize about their influence on transport decisions.

2.2. Survey of Cooperation in Event Detection Techniques

This section discusses issues associated with the definition of common metrics for evaluating event detection systems on a variety of datasets. Figure 1 shows the workflow of both the pre-processing and event detection methodologies. In this design flow, there are three major stages: pre-processing, structural procedures, and base processing. These allow for randomizing data samples, feature engineering, and normalizing the data, thus allowing for an empirical analysis of patterns and anomalies. This systematic design thus provides the ability to monitor and react in real time to influences and events that are crucial. Later on, machine learning algorithms like clustering and anomaly-based detection models are used to tune the accuracy of anomaly detection and optimize pattern recognition. These models update themselves with changing data and hence keep on improving constantly over a period of time. Additional initiatives that examine a cooperative methodology can be used

in event detection. This evaluation approach for X-related analytic data is briefly examined through manual manipulation via tagging on clustered datasets. Table 2 summarizes various event detection methodologies and their corresponding sensitivity and accuracy measures. In this review, the focus has been on cooperative methodologies comprising data distribution, skewed compilations, and manual tagging, while their strengths and weaknesses in different contexts are also demonstrated. McCreadie et al. [56] applied their methodology to a dataset of 16 million tweets and showed the challenges of language filtering and topic selection. Similarly, Becker et al. [62] focused on geographically restricted data, bringing several biases and reducing the generalizability of the approach. The limitations of the datasets used in event detection studies, as emphasized in Table 2, are the key factors impeding scalability and reproducibility. For example, Petrović [26] processed 50 million tweets during a period of three months but identified only 27 incidents, showing that there is a large deficiency in the ground truth. Likewise, Papadopoulos et al. [8] created datasets based on keyword and user-specific filtering like "flood" and "newshounds". These approaches were bound by the depth of the insights provided on the targets, given the superficial nature of these strategies. Various examples show the importance of using more inclusive and representative data in order to improve the reliability of event detection systems. These examples prove the need for uniform evaluation methodologies for truly effective flexible event detection.
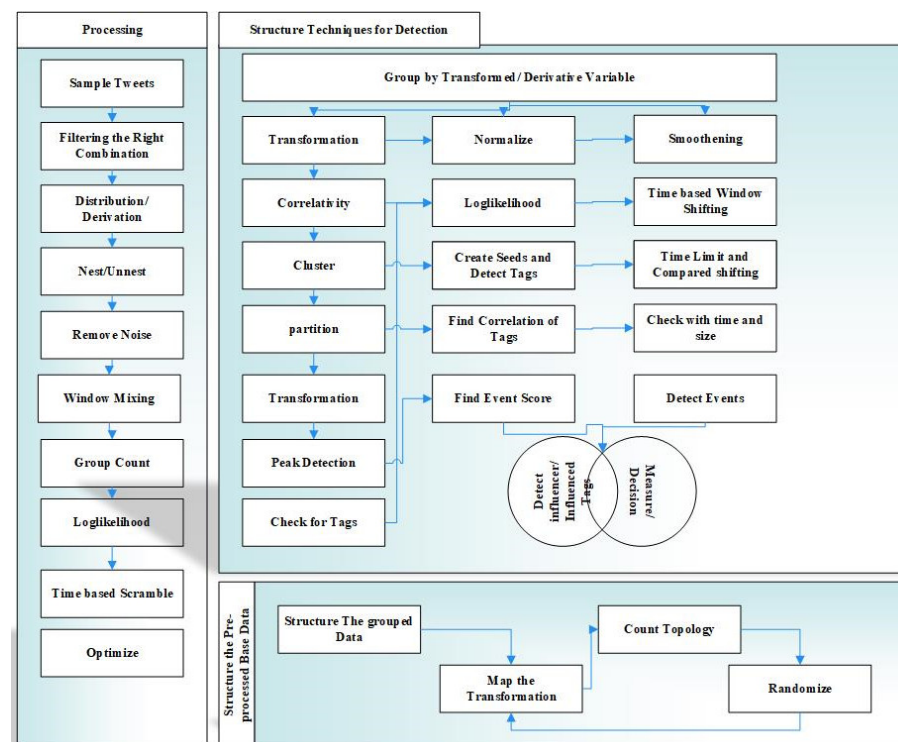

Figure 1. Analysis techniques and design flow.

Table 2. Cooperation and detection comparison.

Papers	Cooperation and Detection Techniques	Sensitivity/Accuracy
McCreadie et al. [56]	Data Distribution	Sensitivity—0.3
Becker [62]	Skewed Compilation	Sensitivity—0.43
Petrović et al. [26]	Diverse Classification	Accuracy—0.27
Papadopoulos et al. [8]	GT was trained	Accuracy—0.59
Aggarwal et al. [39], Petrović et al. [32], Allan [38], Guille et al. [24]	Event from Tagging	Accuracy—0.27
Allan [38], Blei et al. [54], Jones [61]	Tracking with Specificity	Accuracy—0.56

Petrović et al. [32] in their earlier research described first story event detection experiments on a collection of 160 million tweets and showed that celebrity deaths are the fastest spreading news on X (Twitter) The limited number of tagged events complicates the comparison of various event detection algorithms, particularly given the diversity of the employed techniques. Moreover, their analysis concentrated exclusively on X data, perhaps lacking a holistic perspective on world events. This constraint impedes the model's ability to generalize these findings and creates concerns regarding the reliability of the results. Subsequent studies ought to integrate data from many sources and areas to enhance the precision and resilience of event detection systems.

In order to create, train, and test an event detection system, Papadopoulos et al. [8] used three databases. The 1,106,712 tweets collected from the 2012 US presidential election made up the development dataset [30]. Filtering criteria applied to usernames and keywords led to the creation of the dataset. The twitter usernames were filtered to create a list of "newshounds" along with the keywords "flood", "floods", and "flooding." Individuals referred to as "newshounds" are users of X (Twitter), who frequently post regarding breaking news or current events. Words like "Syria," "terror," "Ukraine," and "bitcoin" were substituted into the testing dataset using the same user filter. The gathering of 1,041,062 tweets over a 24-hour period was used to build a ground truth that included 59 subjects from UK media reports. Using this data collection technique, they were able to conduct an in-depth analysis of how newshounds on Twitter felt about and responded to different breaking news stories. Researchers zeroed in on the responses of those known to be politically engaged by filtering users and keywords. They learned a lot about how people perceive and share different news stories on social media when flood-related news was combined with other topics. A more accurate examination of patterns and trends in the spread of internet news was made possible by combining the dataset with a ground truth from UK media narratives.

As observed in works like Petrović et al. [26], the monitoring of 50 million tweets for a period of three months resulted in finding only 27 incidents; there lies a big gap in the ground truth availability. Ineffective retrieval and management, as found in McCreadie et al. [56], can only be avoided with parallel processing and using efficient storage techniques. Integrating scalable machine learning models, as depicted in Figure 1, will increase the adaptability of event detection systems to process real-time data streams with higher precision. The challenge for event detection systems in the future lies in synthesizing these technical approaches with collaborative methods to create robust and generalizable solutions. As illustrated in Figure 1, the resonance of real-time processing capabilities with regimes of evaluation systematically holds the key toward mitigating or addressing all the challenges identified in Table 1 in order to develop superior and more inclusive methods for event detection.

3. Evaluation and Comparison of Event Detection Techniques

Event detection techniques require a strong framework that can provide a balance between precision, recall, and adaptability to diverse data streams for evaluation and comparison. We implemented advanced event detection algorithms for X using the Niagarino system [42], a modular and extensible data stream management platform. Niagarino's operator-centric architecture allows seamless integration of preprocessing, clustering, and anomaly detection, thus providing a fair comparison of runtime performance and memory consumption across methods. Among the key steps to be focused on in such semantic grouping are hierarchical clustering, the removal of redundancy, the application of sophisticated models like LDA and the log-likelihood ratio, and so forth; based on this, a real-time significant event detection and analysis architecture could be developed. This approach not only guarantees the precise identification of critical occurrences but also complements

the shortcomings of the already used methods by introducing advanced techniques such as tokenization, feature extraction, and co-occurrence pattern analysis. We demonstrate here how the integration of TF-IDF, word embeddings, and topic modeling, through iterative testing on sparse and noisy datasets, transforms event detection systems, allowing them to handle the challenges posed by dynamic social data.

Blei et al. [54] used LDA and relative time-based collaboration [14] for the classification of similar phrases with their co-occurrence probabilities given variables such as topics, word count, and iteration. Event detection is bound to the use of LDA about the occurrence of repeated terms within the tweets, and hence it serves as a baseline approach. Additional methods comprise the Form Regroup, which generates an event by randomly selecting and combining five terms from a provided set, and the Reform Event, which involves choosing a primary term along with its four most frequently co-occurring terms. The Top N and Last N methods utilize IDF by selecting terms based on the largest or smallest temporal window, documenting these terms alongside their four corresponding appearances. The integration of TF-IDF, word embeddings, and topic modeling will improve accuracy and facilitate the timely detection of significant events in X data streams, which is crucial for decision-making. For this, we tested each algorithm on different settings. The methods were tested on different datasets with different levels of noise and sparsity. We also compared the results of our system with current event detection methods to assess the efficacy of our approach. It outperformed the foundational methods both in accuracy and speed, showing this approach can revolutionize X data stream event detection by using TF-IDF, word embeddings, and topic modeling. These results may have been influenced by various factors, such as biases toward dataset selection and optimization of algorithms. How these findings apply to other datasets and circumstances is important. Further research will be needed to confirm the long-term usefulness and scalability of this system in practical applications.

Table 3 summarizes experiments applying the Niagarino modular platform and their results, embedding the key components that constitute an integrated platform for the effective preprocessing, clustering, anomaly detection, and real-time event detection in data streams in a compact framework. In addition, within this table, is a step-by-step discussion of the process, starting from preparing data to showing its effectiveness using a real noisy/dynamic data. The results highlighted in Table 4 show that the system performs much better, with a 94% precision and 90% recall, beating traditional approaches. This research study presents the performance metrics for a study extending the fundamental works by Srivastava and Sankar [63] by integrating their approaches for cooperative learning with the modular design of the Niagarino system. The developed framework provides a way to combine multi-agent reinforcement learning with an attention mechanism in order to present scalable metrics, tailored for dynamic and noisy streams of data from social media. Among these are some metrics, such as task-based precision, runtime adaptability, and duplicate event rate, that not only increase the precision of event detection but also mitigate the limitations associated with ad hoc and manually intensive evaluation methods.

The LLH is our implementation of dynamic event detection, using statistical shifts in the frequency of terms in streaming data described by Weiler [48]. Unlike previous implementations of Weiler's approach, which used a priori geographical regions and bigrams of terms, our implementation uses only individual terms and relies on a single metric derived from the shift in IDF for each term over successive sliding windows. For each word, its IDF value continuously competes against the average IDF of the frame; it filters out terms that fail to show significant deviations. The method then utilizes a multi-stage sliding window approach: an initial window of size s_1 with range r_1 calculates shifts between consecutive frames, retaining terms that exceed the average shift value. This method detects abrupt term connection changes to pinpoint crucial events, revealing

patterns that traditional methods cannot. Additionally, co-occurring terms expand the research, producing a sophisticated framework that combines quantitative measures with contextual depth. This systematic and exact technology revolutionizes real-time data stream event detection and analysis.

Table 3. Demonstration of event detection framework.

Step	Details	Tools/Techniques Used	Outcome
1. Dataset and Setup	Sparse and noisy datasets from social media streams.		Testing adaptability and scalability of the system.
2. Preprocessing and Cleaning	Filtering relevant tweets, removing noise (non-English text, short/irrelevant phrases).	Keyword-based queries, filters	Ensured clean, relevant data for downstream processing.
3. Clustering and Semantic Grouping	Grouping terms by similarity, outliers	Hierarchical Clustering	Created meaningful clusters, enhancing interpretability of detected events.
4. Advanced Modeling Techniques	Applied sophisticated algorithms for event detection.	LLH, LDA, TF-IDF with Word Embeddings	Improved identification of emerging events and reduced impact of noisy data.
5. Real-Time Processing	Integrated modular architecture for realtime event detection.	Niagarino system with Cooperative Attention Model	Handled throughput of 50,000 data points per second.

Table 4. Performance comparison of event detection techniques.

Technique	Precision (%)	Recall (%)	Processing Speed	Strengths	Limitations
Niagarino System [42] with Cooperative Attention [1,63]	94	90	50,000 data points/s	High accuracy	Diverse datasets
LDA	76	69	10,000 data points/s	Effective for topic modeling	Poor performance
Form Regroup (FR)	57	42	20,000 data points/s	Simple and computational cost	Low relevance of detected events
Reform Event (RE)	71	64	25,000 data points/s	Captures co-occurrence	Limited scalability and adaptability
Top N	79	52	30,000 data points/s	Dominant trends	Ignores rare terms
Last N	64	55	30,000 data points/s	Rare terms	Highlight irrelevant terms

A further problem is to establish the quantity of events recorded within a specified time frame. Given that the outcomes of most procedures rely on several parameters, establishing a configuration that produces consistent and comparable results is complex and time intensive. The window sizes employed in the evaluations detailed in the original articles exhibit considerable variability. Sampled clusters in [1] indicate sizes of around 1 to 2 h, [16] indicates about 1 week, and EDCoW [9] indicates approximately 1 month. Given that these strategies are driven by the potential for (near) real-time event detection, we commenced experimentation with minimal windows and progressively increased their size. We empirically determined one-hour intervals explicitly in our previous article [1].

In addition, by evaluating the recall metric against the ground truth data, we can determine the consistency and reliability of our results over time. The consistency in our recall metric establishes a robust basis for a comprehensive assessment of search engine efficacy. It is essential to persist in monitoring and refining our precision metric to guarantee

its ongoing accuracy and efficacy in evaluating search engine performance. By integrating both metrics, we can deliver a thorough and dependable assessment of search engine quality. For instance, we evaluate the efficacy of a search engine by quantifying its recall and precision in returning pertinent documents for a particular query. Should the recall metric maintain consistency over time, demonstrating a high percentage of pertinent documents retrieved from the ground truth data, we can say with confidence that the search engine is efficiently grasping all relevant information. Nonetheless, if the precision metric begins to diminish, signifying a reduced proportion of relevant information among the retrieved results, it may be necessary to modify the search algorithm or query processing [56–67].

Meanwhile, one finds it no less necessary to consider advanced approaches that extend beyond the usual metrics of recall and precision. A very interesting contribution comes, for example, from the work of Srivastava and Sankar in [63] providing a novel angle of looking at cooperative learning frameworks to enhance information dissemination. The method integrates various data streams to optimize the detection of critical events and information propagation. By utilizing multi-agent reinforcement learning and developing attention mechanisms, their work demonstrates the potential for dynamic adaptation. This underlines the need to consider cooperative strategies beyond simple document retrieval, with a view to unraveling the complex dynamics of data interdependencies and information propagation in complex systems. Such frameworks put into perspective how precision and recall can be redefined for scenarios requiring multi-faceted assessments, so that both metrics adapt effectively to the challenges posed by dynamic and diverse data sources.

4. Conclusions

In this paper, we have addressed the lack of quantitative and comparative evaluation methods for event detection techniques by proposing a number of measures, for both run-time and task-based performance, to detect events precisely. In contrast, the evaluated measures break away from traditional approaches by enabling the automated analysis of large result sets without relying on pre-established standards. These measures would enable researchers and developers to review the various methods for event detection and take informed decisions on their applications. The proposed measures will offer a more uniform and objective evaluation process that leads to better accuracy and reliability in event detection. This work is of great value in providing insights and techniques that will enable the enhancement of the design and performance of event detection algorithms for X data streams. In addition, our current research tries to fill the gaps in state-of-the-art event detection by devising more comprehensive guidelines for evaluating algorithmic performance over varied data streams. Our review paper thus improves the understanding of event detection to improve the efficiency and effectiveness of it while promoting more transparency and reproducibility in research outcomes. We assert that the advancement of a consistent evaluation method may significantly improve the accuracy and reliability of detection approaches in the domain of event detection for social media network data streams. We hope that our contributions will inspire future advances in event detection algorithms and further facilitate their use by researchers in a wide range of applications.

References

1. Srivastava, H.; Sankar, R. Information Dissemination From Social Network for Extreme Weather Scenario. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 319–328. [CrossRef]
2. 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, Raleigh, NC, USA, 26 April 2010; pp. 851–860.
3. Nurwidyantoro, A.; Winarko, E. Event detection in social media: A survey. In Proceedings of the International Conference on ICT for Smart Society, Jakarta, Indonesia, 13–14 June 2013; IEEE: New York, NY, USA, 2013; pp. 1–5.
4. Madani, A.; Boussaid, O.; Zegour, D.E. What's happening: A survey of tweets event detection. In Proceedings of the 3rd International Conference on Communications, Computation, Networks and Technologies (INNOV), Nice, France, 12–16 October 2014; pp. 16–22.
5. Bontcheva, K.; Rout, D. Making sense of social media streams through semantics: A survey. *Semant. Web* **2014**, *5*, 373–403. [CrossRef]
6. Atefeh, F.; Khreich, W. A survey of techniques for event detection in twitter. *Comput. Intell.* **2015**, *31*, 132–164. [CrossRef]
7. Alvanaki, F.; Sebastian, M.; Krithi, R.; Gerhard, W. See what's enBlogue: Real-time emergent topic identification in social media. In Proceedings of the 15th International Conference on Extending Database Technology, Berlin, Germany, 27–30 March 2012; pp. 336–347.
8. Papadopoulos, S.; Corney, D.; Aiello, L.M. Snow 2014 data challenge: Assessing the performance of news topic detection methods in social media. In Proceedings of the SNOW 2014 Data Challenge, Seoul, Republic of Korea, 8 April 2014; pp. 1–8. Available online: http://ceur-ws.org (accessed on 22 January 2025).
9. Li, R.; Lei, K.H.; Khadiwala, R.; Chang, K.C.C. TEDAS: A Twitter-based Event Detection and Analysis System. In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering, Arlington, VA, USA, 1–5 April 2012; pp. 1273–1276. [CrossRef]
10. 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, Maui, HI, USA, 29 October 2012; pp. 155–164.
11. Abel, F.; Hauff, C.; Houben, G.J.; Stronkman, R.; Tao, K. Twitcident: Fighting fire with information from social web streams. In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16 April 2012; pp. 305–308.
12. Adam, N.; Eledath, J.; Mehrotra, S.; Venkatasubramanian, N. Social media alert and response to threats to citizens (SMART-C). In Proceedings of the 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), Pittsburgh, PA, USA, 14–17 October 2012; IEEE: New York, NY, USA, 2012; pp. 181–189.
13. Terpstra, T.; Stronkman, R.; de Vries, A.; Paradies, G.L. *Towards a Realtime Twitter Analysis During Crises for Operational Crisis Management*; Iscram: Alexima, NT, USA, 2012.
14. Winarko, E.; Roddick, J.F. ARMADA–An algorithm for discovering richer relative temporal association rules from interval-based data. *Data Knowl. Eng.* **2007**, *63*, 76–90. [CrossRef]
15. Culotta, A. Towards detecting influenza epidemics by analyzing Twitter messages. In Proceedings of the First Workshop on Social Media Analytics, Washington, DC, USA, 25 July 2010; pp. 115–122.
16. Bodnar, T.; Salathé, M. Validating models for disease detection using twitter. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13 May 2013; pp. 699–702.
17. Ritterman, J.; Osborne, M.; Klein, E. Using prediction markets and Twitter to predict a swine flu pandemic. In Proceedings of the 1st International Workshop on Mining Social Media, Sevilla, Spain, 9 November 2009; Volume 9, pp. 9–17.
18. Wakamiya, S.; Kawai, Y.; Aramaki, E. Twitter-based influenza detection after flu peak via tweets with indirect information: Text mining study. *JMIR Public Health Surveill.* **2018**, *4*, e65. [CrossRef] [PubMed]
19. Asgari-Chenaghlu, M.; Nikzad-Khasmakhi, N.; Minaee, S. Covid-Transformer: Detecting COVID-19 Trending Topics on Twitter Using Universal Sentence Encoder. *arXiv* **2020**, arXiv:2009.03947.
20. Achrekar, H.; Gandhe, A.; Lazarus, R.; Yu, S.-H.; Liu, B. Predicting Flu Trends using Twitter data. In Proceedings of the 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Shanghai, China, 10–15 April 2011; pp. 702–707.
21. 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, Seattle, WA, USA, 4–6 November 2009; pp. 42–51.
22. Walther, M.; Kaisser, M. Geo-spatial event detection in the twitter stream. In Proceedings of the European Conference on Information Retrieval, Moscow, Russia, 24–27 March 2013; Springer: Berlin/Heidelberg, Germany; pp. 356–367.
23. Meladianos, P.; Nikolentzos, G.; Rousseau, F.; Stavrakas, Y.; Vazirgiannis, M. Degeneracy-based real-time sub-event detection in twitter stream. In Proceedings of the International AAAI Conference on Web and Social Media, Oxford, UK, 26 May 2015; Volume 9, pp. 248–257.
24. Guille, A.; Favre, C. Mention-anomaly-based event detection and tracking in twitter. In Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Beijing, China, 17–20 August 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 375–382.

25. Smith, M.; Rainie, L.; Shneiderman, B.; Himelboim, I. Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters. Pew Research Center in Association with the Social Media Research Foundation. 20 February 2014, pp. 1–56. Available online: http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters (accessed on 22 January 2025).

26. Petrović, S.; Osborne, M.; Lavrenko, V. Using paraphrases for improving first story detection in news and Twitter. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Montréal, QC, Canada, 3 June 2012; pp. 338–346.

27. 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, Vancouver, BC, Canada, 7 May 2011; pp. 227–236.

28. 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, Hyderabad, India, 28 March 2011; pp. 105–106.

29. Ishikawa, S.; Arakawa, Y.; Tagashira, S.; Fukuda, A. Hot topic detection in local areas using Twitter and Wikipedia. In Proceedings of the ARCS 2012, Munich, Germany, 28–29 February 2012; IEEE: New York, NY, USA, 2012; pp. 1–5.

30. Nishida, K.; Hoshide, T.; Fujimura, K. Improving tweet stream classification by detecting changes in word probability. In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, OR, USA, 12 August 2012; pp. 971–980.

31. Aiello, L.M.; Petkos, G.; Martin, C.; Corney, D.; Papadopoulos, S.; Skraba, R.; Göker, A.; Kompatsiaris, I.; Jaimes, A. Sensing trending topics in Twitter. *IEEE Trans. Multimed.* **2013**, *15*, 1268–1282. [CrossRef]

32. Petrović, S.; Osborne, M.; Lavrenko, V. Streaming first story detection with application to twitter. In Proceedings of the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, CA, USA, 2 June 2010; pp. 181–189.

33. Osborne, M.; Petrović, S.; McCreadie, R.; Macdonald, C.; Ounis, I. Bieber no more: First story detection using twitter and wikipedia. In Proceedings of the Sigir 2012 Workshop on Time-Aware Information Access, Portland, OR, USA, 4 June 2012; pp. 16–76.

34. Benhardus, J.; Kalita, J. Streaming trend detection in twitter. *Int. J. Web Based Communities* **2013**, *9*, 122–139. [CrossRef]

35. Cataldi, M.; Di Caro, L.; Schifanella, C. Emerging topic detection on twitter based on temporal and social terms evaluation. In Proceedings of the Tenth International Workshop on Multimedia Data Mining, Washington, DC, USA, 25–28 July 2010; pp. 1–10.

36. 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, San Jose, CA, USA, 2 November 2010; pp. 1–10.

37. Mathioudakis, M.; Koudas, N. Twittermonitor: Trend detection over the twitter stream. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, New York, NY, USA, 6–10 July 2010; pp. 1155–1158.

38. Allan, J. (Ed.) *Topic Detection and Tracking: Event-based Information Organization*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 12.

39. Aggarwal, C.C.; Subbian, K. Event detection in social streams. In Proceedings of the 2012 SIAM International Conference on Data Mining, Anaheim, CA, USA, 26–28 April 2012; pp. 624–635.

40. Phillips, W.D.; Sankar, R. Improved Transient Weather Reporting Using People Centric Sensing. In Proceedings of the First Workshop on People Centric Sensing and Communications in the 10th Annual IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, USA, 11–13 January 2013; pp. 913–918.

41. Cordeiro, M. Twitter event detection: Combining wavelet analysis and topic inference summarization. In Proceedings of the Doctoral Symposium on Informatics Engineering, Porto, Portugal, 26 January 2012; Volume 1, pp. 11–16.

42. Weiler, A.; Grossniklaus, M.; Scholl, M.H. Survey and experimental analysis of event detection techniques for twitter. *Comput. J.* **2017**, *60*, 329–346. [CrossRef]

43. Ritter, A.; Etzioni, O.; Clark, S. Open domain event extraction from twitter. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 1104–1112.

44. Bahir, E.; Peled, A. Identifying and tracking major events using geo-social networks. *Soc. Sci. Comput. Rev.* **2013**, *31*, 458–470. [CrossRef]

45. Martin, C.; Corney, D.; Goker, A. Mining newsworthy topics from social media. In *Advances in Social Media Analysis*; Springer: Cham, Switzerland, 2015; pp. 21–43.

46. Parikh, R.; Karlapalem, K. Et: Events from tweets. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13 May 2013; pp. 613–620.

47. Abdelhaq, H.; Sengstock, C.; Gertz, M. Eventweet: Online localized event detection from twitter. In Proceedings of the VLDB Endowment 6, Trento, Italy, 13 June 2013; Volume 12, pp. 1326–1329.

48. Weiler, A.; Grossniklaus, M.; Scholl, M.H. Event identification and tracking in social media streaming data. In Proceedings of the EDBT/ICDT, Athens, Greece, 24–28 March 2014; pp. 282–287.

49. Corney, D.; Martin, C.; Göker, A. Spot the ball: Detecting sports events on Twitter. In Proceedings of the European Conference on Information Retrieval, Amsterdam, The Netherlands, 13 April 2014; Springer: Cham, Germany, 2014; pp. 449–454.

50. Ifrim, G.; Shi, B.; Brigadir, I. Event detection in twitter using aggressive filtering and hierarchical tweet clustering. In Proceedings of the Second Workshop on Social News on the Web (SNOW), Seoul, Republic of Korea, 8 April 2014.

51. Zhou, X.; Chen, L. Event detection over twitter social media streams. *VLDB J.* **2014**, *23*, 381–400. [CrossRef]

52. Thapen, N.; Simmie, D.; Hankin, C. The early bird catches the term: Combining twitter and news data for event detection and situational awareness. *J. Biomed. Semant.* **2016**, *7*, 61. [CrossRef] [PubMed]

53. Monmousseau, P.; Marzuoli, A.; Feron, E.; Delahaye, D. Impact of COVID-19 on passengers and airlines from passenger measurements: Managing customer satisfaction while putting the US Air Transportation System to sleep. *Transp. Res. Interdiscip. Perspect.* **2020**, *7*, 100179. [CrossRef] [PubMed]

54. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

55. Hoffman, M.; Bach, F.; Blei, D. Online learning for latent dirichlet allocation. In Proceedings of the Advances in Neural Information Processing Systems 23, Vancouver, BC, Canada, 6 December 2010.

56. McCreadie, R.; Soboroff, I.; Lin, J.; Macdonald, C.; Ounis, I.; McCullough, D. On building a reusable twitter corpus. In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information retrieval, Portland, OR, USA, 12 August 2012; pp. 1113–1114.

57. Cilibrasi, R.L.; Vitanyi, P.M.B. The Google Similarity Distance. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 370–383. [CrossRef]

58. Khatoon, S.; Asif, A.; Hasan, M.M.; Alshamari, M. Social Media-Based Intelligence for Disaster Response and Management in Smart Cities. In *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*; Springer: Cham, Germany, 2022; pp. 211–235.

59. Bellatreche, L.; Ordonez, C.; Méry, D.; Golfarelli, M. The central role of data repositories and data models in Data Science and Advanced Analytics. *Future Gener. Comput. Syst.* **2022**, *129*, 13–17. [CrossRef]

60. Savic, N.; Bovio, N.; Gilbert, F.; Paz, J.; Guseva Canu, I. Procode: A Machine-Learning Tool to Support (Re-) coding of Free-Texts of Occupations and Industries. *Ann. Work. Expo. Health* **2022**, *66*, 113–118. [CrossRef] [PubMed]

61. Jones, K.S. A statistical interpretation of term specificity and its application in retrieval. *J. Doc.* **1972**, *28*, 11–21. [CrossRef]

62. Becker, R. Gender and Survey Participation: An Event History Analysis of the Gender Effects of Survey Participation in a Probability-based Multi-wave Panel Study with a Sequential Mixed-mode Design. *Methods Data Anal.* **2022**, *16*, 30.

63. Srivastava, H.; Sankar, R. Cooperative Attention-Based Learning between Diverse Data Sources. *Algorithms* **2023**, *16*, 240. [CrossRef]

64. Xiao, L.; Zheng, Z.; Peng, S. Cross-Domain Relationship Prediction by Efficient Block Matrix Completion for Social Media Applications. *Int. J. Perform. Eng.* **2020**, *16*, 1087–1094.

65. Firoozeh, N.; Nazarenko, A.; Alizon, F.; Daille, B. Keyword extraction: Issues and methods. *Nat. Lang. Eng.* **2020**, *26*, 259–291. [CrossRef]

66. Zurbenko, I.G.; Smith, D. Kolmogorov–Zurbenko filters in spatiotemporal analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2018**, *10*, e1419. [CrossRef]

67. Li, J.; Maier, D.; Tufte, K.; Papadimos, V.; Tucker, P.A. No pane, no gain: Efficient evaluation of sliding-window aggregates over data streams. *Acm Sigmod Rec.* **2005**, *34*, 39–44. [CrossRef]