arXiv:2007.09815v3 [cs.AI] 4 Aug 2020

Event Prediction in the Big Data Era: A Systematic Survey

LIANG ZHAO, Emory University

Events are occurrences in specific locations, time, and semantics that nontrivially impact either our society or the nature, such as earthquakes, civil unrest, system failures, pandemics, and crimes. It is highly desirable to be able to anticipate the occurrence of such events in advance in order to reduce the potential social upheaval and damage caused. Event prediction, which has traditionally been prohibitively challenging, is now becoming a viable option in the big data era and is thus experiencing rapid growth, also thanks to advances in high performance computers and new Artificial Intelligence techniques. There is a large amount of existing work that focuses on addressing the challenges involved, including heterogeneous multi-faceted outputs, complex (e.g., spatial, temporal, and semantic) dependencies, and streaming data feeds. Due to the strong interdisciplinary nature of event prediction problems, most existing event prediction methods were initially designed to deal with specific application domains, though the techniques and evaluation procedures utilized are usually generalizable across different domains. However, it is imperative yet difficult to cross-reference the techniques across different domains, given the absence of a comprehensive literature survey for event prediction. This paper aims to provide a systematic and comprehensive survey of the technologies, applications, and evaluations of event prediction in the big data era. First, systematic categorization and summary of existing techniques are presented, which facilitate domain experts' searches for suitable techniques and help model developers consolidate their research at the frontiers. Then, comprehensive categorization and summary of major application domains are provided to introduce wider applications to model developers to help them expand the impacts of their research. Evaluation metrics and procedures are summarized and standardized to unify the understanding of model performance among stakeholders, model developers, and domain experts in various application domains. Finally, open problems and future directions for this promising and important domain are elucidated and discussed.

Additional Key Words and Phrases: Event Prediction, Big Data, Artificial Intelligence

ACM Reference Format:
Liang Zhao. 2020. Event Prediction in the Big Data Era: A Systematic Survey. 1, 1 (August 2020), 40 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

An event is a real-world occurrence that takes place in a specific location and time that relates to a particular topic. Events can range from large-scale (e.g., civil unrest events and earthquakes), to medium-scale (e.g., system failures and crime incidents), to small-scale (e.g., authentication events and individual actions) occurrences [10]. Event analytics are important in domains as different as healthcare, business, cybersphere, politics, and entertainment, influencing almost every corner of our lives [222]. The analysis of events has thus been attracting huge attention over the last few decades and can be categorized in terms of their timeliness for various research directions, such as event summarization, detection, and prediction. Unlike retrospective analyses such as event summarization and detection [38], event prediction focuses on anticipating events in the future

Author's address: Liang Zhao, liang.zhao@emory.edu, Emory University.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2020 Association for Computing Machinery.
XXXX-XXXX/2020/8-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

and is the focus of this survey. Accurate anticipation of future events enables one to maximize the benefits and minimize the losses associated with some event in the future, bringing huge benefits for both society as a whole and individual members of society in key domains such as disease prevention [167], disaster management [140], business intelligence [226], and economics stability [24].

"Prediction is very difficult, especially if it's about the future." – Niels Bohr, 1970

Event prediction has traditionally been prohibitively challenging across different domains, due to the lack or incompleteness of our knowledge regarding the true causes and mechanisms driving event occurrences in most domains. With the advent of the big data era, however, we now enjoy unprecedented opportunities that open up many alternative approaches for dealing with event prediction problems, sidestepping the need to develop a complete understanding of the underlying mechanisms of event occurrence. Based on large amounts of data on historical events and their potential precursors, event prediction methods typically strive to apply predictive mapping to build on these observations to predict future events, utilizing predictive analysis techniques from domains such as machine learning, data mining, pattern recognition, statistics, and other computational models [16, 26, 92]. Event prediction is currently experiencing extremely rapid growth, thanks to advances in sensing techniques (physical sensors and social sensors), prediction techniques (Artificial Intelligence, especially Machine Learning), and high performance computing hardware [78].

Event prediction in big data is a difficult problem that requires the invention and integration of related techniques to address the serious challenges caused by its unique characteristics, including: **1) Heterogeneous multi-output predictions.** Event prediction methods usually need to predict multiple facets of events including their time, location, topic, intensity, and duration, each of which may utilize a different data structure [171]. This creates unique challenges, including how to jointly predict these heterogeneous yet correlated facets of outputs. Due to the rich information in the outputs, label preparation is usually a highly labor-intensive task performed by human annotators, with automatic methods introducing numerous errors in items such as event coding. So, how can we improve the label quality as well as the model robustness under corrupted labels? The multi-faceted nature of events make event prediction a multi-objective problem, which raises the question of how to properly unify the prediction performance on different facets. It is also challenging to verify whether a predicted event "matches" a real event, given that the various facets are seldom, if ever, 100% accurately predicted. So, how can we set up the criteria needed to discriminate between a correct prediction ("true positive") and a wrong one ("false positive")? **2) Complex dependencies among the prediction outputs.** Beyond conventional isolated tasks in machine learning and predictive analysis, in event prediction the predicted events can correlate to and influence each other [142]. For example, an ongoing traffic incident event could cause congestion on the current road segment in the first 5 minutes but then lead to congestion on other contiguous road segments 10 minutes later. Global climate data might indicate a drought in one location, which could then cause famine in the area and lead to a mass exodus of refugees moving to another location. So, how should we consider the correlations among future events? **3) Real-time stream of prediction tasks.** Event prediction usually requires continuous monitoring of the observed input data in order to trigger timely alerts of future potential events [182]. However, during this process the trained prediction model gradually becomes outdated, as real world events continually change dynamically, concepts are fluid and distribution drifts are inevitable. For example, in September 2008 21% of the United States population were social media users, including 2% of those over 65. However, by May 2018, 72% of the United States population were social media users, including 40% of those over [37]. Not only the data distribution, but also the number of features and input data

sources can also vary in real time. Hence, it is imperative to periodically upgrade the models, which raises further questions concerning how to train models based on non-stationary distributions, while balancing the cost (such as computation cost and data annotation cost) and timeliness? In addition, event prediction involves many other common yet open challenges, such as imbalanced data (for example data that lacks positive labels in rare event prediction) [206], data corruption in inputs [248], the uncertainty of predictions [25], longer-term predictions (including how to trade-off prediction accuracy and lead time) [27], trade-offs between precision and recall [171], and how to deal with high-dimensionality [247] and sparse data involving many unrelated features [208]. Event prediction problems provide unique testbeds for jointly handling such challenges.

In recent years, a considerable amount of research has been devoted to event prediction technique development and applications, in order to address the aforementioned challenges [157]. Recently, there has been a surge of research that both proposes and applies new approaches in numerous domains, though event prediction techniques are generally still in their infancy. Most existing event prediction methods have been designed for a specific application domains, but their approaches are usually general enough to handle problems in other application domains. Unfortunately, it is difficult to cross-reference these techniques across different application domains serving totally different communities. Moreover, the quality of event prediction results require sophisticated and specially-designed evaluation strategies due to the subject matter's unique characteristics, for example its multi-objective nature (e.g., accuracy, resolution, efficiency, and lead time) and heterogeneous prediction results (e.g., heterogeneity and multi-output). As yet, however, we lack a systematic standardization and comprehensive summarization approaches with which to evaluate the various event prediction methodologies that have been proposed. This absence of a systematic summary and taxonomy of existing techniques and applications in event prediction causes major problems for those working in the field who lacks clear information on the existing bottlenecks, traps, open problems, and potentially fruitful future research directions.

To overcome these hurdles and facilitate the development of better event prediction methodologies and applications, this survey paper aims to provide a comprehensive and systematic review of the current state of the art for event prediction in the big data era. The paper's major contributions include:

- **A systematic categorization and summarization of existing techniques.** Existing event prediction methods are categorized according to their event aspects (time, location, and semantics), problem formulation, and corresponding techniques to create the taxonomy of a generic framework. Relationships, advantages, and disadvantages among different subcategories are discussed, along with details of the techniques under each subcategory. The proposed taxonomy is designed to help domain experts locate the most useful techniques for their targeted problem settings.
- **A comprehensive categorization and summarization of major application domains.** The first taxonomy of event prediction application domains is provided. The practical significance and problem formulation are elucidated for each application domain or subdomain, enabling it to be easily mapped to the proposed technique taxonomy. This will help data scientists and model developers to search for additional application domains and datasets that they can use to evaluate their newly proposed methods, and at the same time expand their advanced techniques to encompass new application domains.
- **Standardized evaluation metrics and procedures.** Due to the nontrivial structure of event prediction outputs, which can contain multiple fields such as time, location, intensity, duration, and topic, this paper proposes a set of standard metrics with which to standardize existing ways to pair predicted events with true events. Then additional metrics are introduced and standardized

to evaluate the overall accuracy and quality of the predictions to assess how close the predicted events are to the real ones.

- **An insightful discussion of the current status of research in this area and future trends.** Based on the comprehensive and systematic survey and investigation of existing event prediction techniques and applications presented here, an overall picture and the shape of the current research frontiers are outlined. The paper concludes by presenting fresh insights into the bottleneck, traps, and open problems, as well as a discussion of possible future directions.

1.1 Related Surveys

This section briefly outlines previous surveys in various domains that have some relevance to event prediction in big data in three categories, namely: 1. event detection, 2. predictive analytics, and 3. domain-specific event prediction.

Event detection has been an extensively explored domain with over many years. Its main purpose is to detect historical or ongoing events rather than to predict as yet unseen events in the future [181, 222]. Event detection typically focuses on pattern recognition [26], anomaly detection [92], and clustering [92], which are very different from those in event prediction. There have been several surveys of research in this domain in the last decade [9, 15, 63, 146]. For example, Deng et al. [63] and Atefeh and Khreich [15] provided overviews of event extraction techniques in social media, while Michelioudakis et al. [146] presented a survey of event recognition with uncerntainty. Alevizos et al. [9] provided a comprehensive literature review of event recognition methods using probabilistic methods.

Predictive analysis covers the prediction of target variables given a set of dependent variables. These target variables are typically homogeneous scalar or vector data for describing items such as economic indices, housing prices, or sentiments. The target variables may not necessarily be values in the future. Larose [120] provides a good tutorial and survey for this domain. Predictive analysis can be broken down into subdomains such as structured prediction [26], spatial prediction [105], and sequence prediction [88], enabling users to handle different types of structure for the target variable. Fülöp et al. [81] provided a survey and categorization of applications that utilize predictive analytics techniques to perform event processing and detection, while Jiang [105] focused on spatial prediction methods that predict the indices that have spatial dependency. Baklr et al. [17] summarized the literature on predicting structural data such as geometric objects and networks, and Arias et al. [12] Phillips et al. [163], and Yu and Kak [232] all proposed the techniques for predictive analysis using social data.

As event prediction methods are typically motivated by specific application domains, there are a number of surveys event predictions for domains such as flood events [47], social unrest [29], wind power ramp forecasting [76], tornado events [68], temporal events without location information [87], online failures [182], and business failures [6]. However, in spite of its promise and its rapid growth in recent years, the domain of event prediction in big data still suffers from the lack of a comprehensive and systematic literature survey covering all its various aspects, including relevant techniques, applications, evaluations, and open problems.

1.2 Outline

The remainder of this article is organized as follows. Section 2 presents generic problem formulations for event prediction and the evaluation of event prediction results. Section 3 then presents a taxonomy and comprehensive description of event prediction techniques, after which Section 4 categorizes and summarizes the various applications of event prediction. Section 5 lists the open problems and suggests future research directions and this survey concludes with a brief summary in Section 6.

2 PROBLEM FORMULATION AND PERFORMANCE EVALUATIONS

This section begins by examining the generic denotation and formulation of the event prediction problem (Section 2.1) and then considers way to standardize event prediction evaluations (Section 2.2).

2.1 Problem Formulation

An event refers to a real-world occurrence that happens at some specific time and location with specific semantic topic [222]. We can use $y = (t, l, s)$ to denote an event where its time $t \in \mathcal{T}$, its location $l \in \mathcal{L}$, and its semantic meaning $s \in \mathcal{S}$. Here, $\mathcal{T}$, $\mathcal{L}$, and $\mathcal{S}$ represent the time domain, location domain, and semantic domain, respectively. Notice that these domains need to have very general meanings that cover a wide range of types of entities. For example, the location $\mathcal{L}$ can include any features that can be used to locate the place of an event in terms of a point or a neighborhood in either Euclidean space (e.g., coordinate and geospatial region) or non-Euclidean space (e.g., a vertex or subgraph in a network). Similarly, the semantic domain $\mathcal{S}$ can contain any type of semantic features that are useful when elaborating the semantics of an event's various aspects, including its actors, objects, actions, magnitude, textual descriptions, and other profiling information. For example, *("11am, June 1, 2019", "Hermosillo, Sonora, Mexico", "Student Protests")* and *("June 1, 2010", âĂIJBerlin, Germany", "Red Cross helps pandemics control")* denote the time, location, and semantics, for two events respectively.

An event prediction system requires inputs that could indicate future events, called event indicators, and these could contain both critical information on events that precede the future event, known as precursors, as well as irrelevant information [86, 171]. Event indicator data can be denoted as $X \subseteq \mathcal{T} \times \mathcal{L} \times \mathcal{F}$, where $\mathcal{F}$ is the domain of the features other than location and time. If we denote the current time as t_{now} and define the past time and future time as $\mathcal{T}^- \equiv \{t | t \leq t_{\text{now}}, t \in \mathcal{T}\}$ and $\mathcal{T}^+ \equiv \{t | t > t_{\text{now}}, t \in \mathcal{T}\}$, respectively, the event prediction problem can now be formulated as follows:

Definition 2.1 (Event Prediction). Given the event indicator data $X \subseteq \mathcal{T}^- \times \mathcal{L} \times \mathcal{F}$ and historical event data $Y_0 \subseteq \mathcal{T}^- \times \mathcal{L} \times \mathcal{S}$, event prediction is a process that outputs a set of predicted future events $\hat{Y} \subseteq \mathcal{T}^+ \times \mathcal{L} \times \mathcal{S}$, such that for each predicted future event $\hat{y} = (t, l, s) \in \hat{Y}$ where $t > t_{\text{now}}$.

Not every event prediction method necessarily focuses on predicting all three domains of time, location, and semantics simultaneously, but may instead predict any part of them. For example, when predicting a clinical event such as the recurrence of disease in a patient, the event location might not always be meaningful [167], but when predicting outbreaks of seasonal flu, the semantic meaning is already known and the focus is the location and time [27] and when predicting political events, sometimes the location, time, and semantics (e.g., event type, participant population type, and event scale) are all necessary [171]. Moreover, due to the intrinsic nature of time, location, and semantic data, the prediction techniques and evaluation metrics of them are necessarily different, as described in the following.

2.2 Event Prediction Evaluation

Event Prediction Evaluation essentially investigates the goodness of fit for a set of predicted events $\hat{Y}$ against real events Y. Unlike the outputs of conventional machine learning models such as the simple scalar values used to indicate class types in classification or numerical values in regression, the outputs of event prediction are entities with rich information. Before we evaluate the quality of prediction, we need to first determine the pairs of predictions and the labels that will be used for the comparison. Hence, we must first optimize the process of matching predictions and real events (Section 2.2.1) before evaluating the prediction error and accuracy (Section 2.2.2).

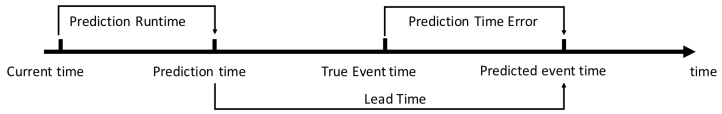

Fig. 1. Various types of times in event prediction.

2.2.1 Matching predicted events and real events. The following two types of matching are typically used:

- **Prefixed matching:** The predicted events will be matched with the corresponding ground-true real events if they share some key attributes. For example, for event prediction at a particular time and location point, we can evaluate the prediction against the ground truth for that time and location. This type of matching is most common when each of the prediction results can be uniquely distinguished along the predefined attributes (for example, location and time) that have a limited number of possible values, so that one-on-one matching between the predicted and real events are easily achieved [1, 244]. For example, to evaluate the quality of a predicted event on June 1, 2019 in San Francisco, USA, the true event occurrence on that date in San Francisco can be used for the evaluation.
- **Optimized matching:** In situations where one-on-one matching is not easily achieved for any event attribute, then the set of predicted events might need to assess the quality of the match achieved with the set of real events, via an optimized matching strategy [168, 171]. For example, consider two predictions, **Prediction 1:** ("9am, June 4, 2019", "Nogales, Sonora, Mexico", "Worker Strike"), and **Prediction 2**: ("11am, June 1, 2019", "Hermosillo, Sonora, Mexico", "Student Protests"). The two ground truth events that these can usefully be compared with are **Real Event 1**: ("9am, June 1, 2019", "Hermosillo, Sonora, Mexico", "Teacher Protests"), and **Real Event 2**: ("June 4, 2019", "Navojoa, Sonora, Mexico", "General-population Protest"). None of the predictions are an exact match for any of the attributes of the real events, so we will need to find a "best" matching among them, which in this case is between **Prediction 1** and **Real Event 2** and **Prediction 2** and **Real Event 1**. This type of matching allows some degree of inaccuracy in the matching process by quantifying the distance between the predicted and real events among all the attribute dimensions. The distance metrics are typically either Euclidean distance [105] or some other distance metric [92]. Some researchers have hired referees to manually check the similarity of semantic meanings [168], but another way is to use event coding to code the events into an event type taxonomy and then consider a match to have been achieved if the event type matches [49].
  Based on the distance between each pair of predicted and real events, the optimal matching will be the one that results in the smallest average distance [157]. However, suppose there are m predicted events and n real events, then there can be as many as $2^{m \cdot n}$ possible ways of matching, making it prohibitively difficult to find the optimal solution. Moreover, there could be different rules for matching. For example, the "multiple-to-multiple" rule shown in Figure 2(a) allows one predicted (real) event to match multiple real (predicted) events [170], while "Bipartite matching" only allows one-to-one matching between predicted and real events (Figure 2(b)). "Non-crossing matching" requires that the real events matched by the predicted events follow the same chronological order (Figure 2(c)). In order to utilize any of these types of matching, researchers have suggested using event matching optimization to learn the optimal set of "(real event, predicted event)" pairs [151].

2.2.2 Metrics of Effectiveness. The effectiveness of the event predictions is evaluated in terms of two indicators: 1) Goodness of Matching, which evaluates performance metrics such as the number

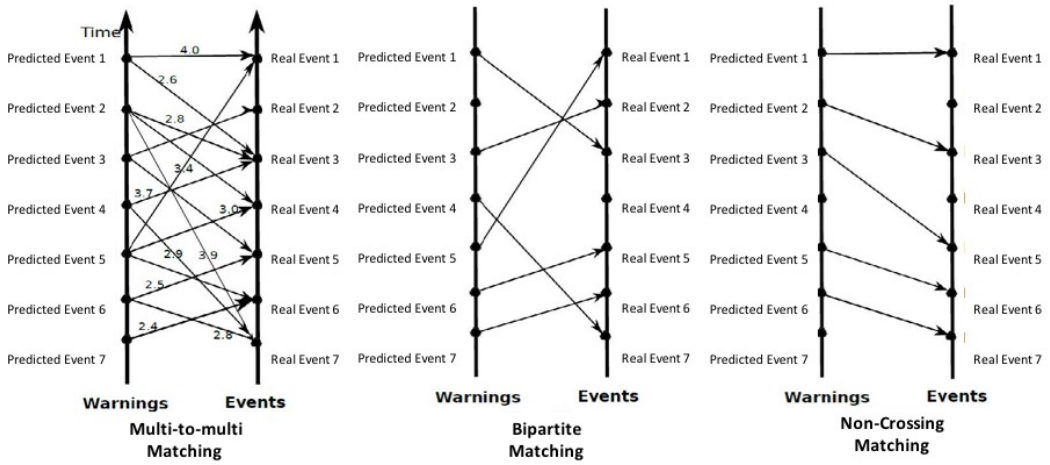

Fig. 2. Generic framework for hierarchical RNN-based event forecasting.

and percentage of matched events [26], and 2) Quality of Matched Predictions, which evaluates how close the predicted event is to the real event for each pair of matched events [171].

- **Goodness of Matching.** A *true positive* means a real event has been successfully matched by a predicted event; if a real event has not been matched by any predicted event, then it is called a *false negative* and a *false positive* means a predicted event has failed to match any real event, which is referred to as a *false alarm*. Assume the total number of predictions is N, the number of real events is $\hat{N}$, the number of true positives is N_{TP}, the number of false negatives is N_{FN} and the number of false positives is N_{FP}. Then, the following key evaluation metrics can be calculated: Prediction=$N_{TP}/(N_{TP} + N_{FP})$, Recall=$N_{TP}/(N_{TP} + N_{FN})$, F-measure = $2 \cdot \text{Precision} \cdot \text{Recall}/(\text{Precision} + \text{Recall})$. Other measurements such as the area under the ROC curves are also commonly used [26]. This approach can be extended to include other items such as multi-class precision/recall, and Precision/Recall at Top K [2, 103, 123, 252].
- **Quality of Matched Predictions.** If a predicted event matches a real one, it is common to go on to evaluate how close they are. This reflects the quality of the matched predictions, in terms of different aspects of the events. Event time is typically a numerical values and hence can be easily measured in terms of metrics such as mean squared error, root mean squared error, and mean absolute error [26]. This is also the case for location in Euclidean space, which can be measured in terms of the Euclidean distance between the predicted point (or region) and the real point (or region). Some researchers consider the administrative unit resolution. For example, a predicted location ("New York City", "New York State", "USA") has a distance of 2 from the real location ("Los Angeles", "California", "USA") [240]. Others prefer to measure multi-resolution location prediction quality as follows: $(1/3)(l_{country} + l_{country} \cdot l_{state} + l_{country} \cdot l_{state} \cdot l_{city})$, where l_{city}, l_{state}, and $l_{country}$ can only be either 0 (i.e., no match to the truth) or 1 (i.e., completely matches the truth) [171]. For a location in non-Euclidean space such as a network [187], the quality can be measured in terms of the shortest path length between the predicted node (or subgraph) and the real node (or subgraph), or by the F-measure between the detected subsets of nodes against the real ones, which is similar to the approach for evaluating community detection [92]. For event topics, in addition to conventional ways of evaluating continuous values such as population size, ordinal values such as event scale, and categorical values such as event type, actors, and actions, as well as more complex semantic values such as texts, can be evaluated using Natural

Language Process measurements such as edit distance, BLEU score, Top-K precision, and ROUGE [10].

3 EVENT PREDICTION TECHNIQUES

This section focuses on the taxonomy and representative techniques utilized for each category and subcategory. Due to the heterogeneity of the prediction output, the technique types depend on the type of output to be predicted, such as time, location, and semantics. As shown in Figure 3, all the event prediction methods are classified in terms of their goals, including time, location, semantics, and the various combinations of these three. These are then further categorized in terms of the output forms of the goals of each and the corresponding techniques normally used, as elaborated in the following.

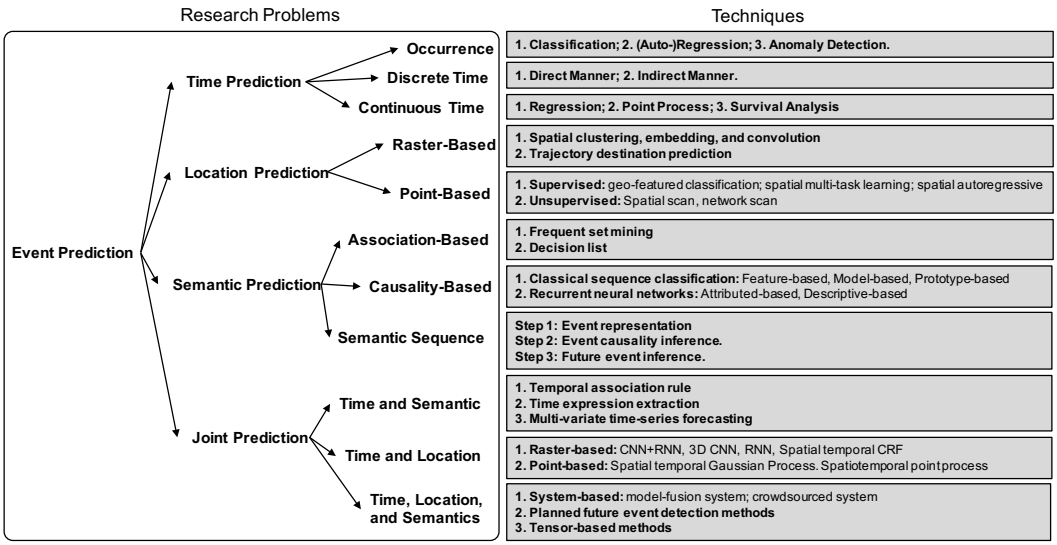

Fig. 3. Taxonomy of event prediction problems and techniques.

3.1 Time Prediction

Event time prediction focuses on predicting when future events will occur. Based on their time granularity, time prediction methods can be categorized into three types: 1) event Occurrence: Binary-valued prediction on whether an event does or does not occur in a future time period; 2) discrete-time prediction: in which future time slot will the event occur; and 3) continuous-time prediction: at which precise time point will the future event occur.

3.1.1 Occurrence Prediction. Occurrence prediction is arguably the most extensive, classical, and generally simplest type of event time prediction task [12]. It focuses on identifying whether there will be event occurrence (positive class) or not (negative class) in a future time period [244]. This problem is usually formulated as a binary classification problem, although a handful of other methods instead leverage anomaly detection or regression-based techniques.

1. Binary classification. Binary classification methods have been extensively explored for event occurrence prediction. The goal here is essentially to estimate and compare the values of $f(y = \text{“}Yes''|X)$ and $f(y = \text{“}No''|X)$, where the former denotes the score or likelihood of event occurrence given observation X while the latter corresponds to no event occurrence. If the value of the former is larger than the latter, then a future event occurrence is predicted, but if not, there is no event

predicted. To implement f, the methods typically used rely on discriminative models, where dedicated feature engineering is leveraged to manually extract potential event precursor features to feed into the models. Over the years, researchers have leveraged various binary classification techniques ranging from the simplest threshold-based methods [121, 251], to more sophisticated methods such as logistic regression [18, 248], Support Vector Machines [102], (Convolutional) Neural Networks [35, 135], and decision trees [58, 185]. In addition to discrminative models, generative models [11, 239] have also been used to embed human knowledge for classifying event occurrences using Bayesian decision techniques. Specifically, instead of assuming that the input features are independent, prior knowledge can also be directly leveraged to establish Bayesian networks among the observed features and variables based on graphical models such as (semi-)hidden Markov models [54, 183, 239] and autoregresive logit models [199]. The joint probabilities $p(y = \text{“}Yes''}, X)$ of $p(y = \text{“}No''}, X)$ can thus be estimated using graphical models, and then utilized to estimate $f(y = \text{“}Yes''}|X) = p(y = \text{“}Yes''}|X)$ and $f(y = \text{“}No''}|X) = p(y = \text{“}No''}|X)$ using Bayesian rules [26].

2. Anomaly detection. Alternatively, anomaly detection can also be utilized to learn a “prototype” of normal samples (typical values corresponding to the situation of no event occurrence), and then identify if any newly-arriving sample is close to or distant from the normal samples, with distant ones being identified as future event occurrences. Such methods are typically utilized to handle “rare event” occurrences, especially when the training data is highly imbalanced with little to no data for “positive” samples. Anomaly detection techniques such as one-classification [189] and hypotheses testing [100, 187] are often utilized here.

3. Regression. In addition to simply predicting the occurrence or not, some researchers have sought to extend the binary prediction problem to deal with ordinal and numerical prediction problems, including event count prediction based on (auto)regression [73], event size prediction using linear regression [229], and event scale prediction using ordinal regression [84].

3.1.2 Discrete-time Prediction. In many applications, practitioners want to know the approximate time (i.e. the date, week, or month) of future events in addition to just their occurrence. To do this, the time is typically first partitioned into different time slots and the various methods focus on identifying which time slot future events are likely to occur in. Existing research on this problem can be classified into either direct or indirect approaches.

1. Direct Approaches. These of methods discretize the future time into discrete values, which can take the form of some number of time windows or time scales such as near future, medium future, or distant future. These are then used to directly predict the integer-valued index of future time windows of the event occurrence using (auto)regression methods [147, 154], or to predict the ordinal values of future time scales using ordinal regression or classification [201].

2. Indirect Approaches. These methods adopt a two-step approach, with the first step being to place the data into a series of time bins and then perform time series forecasting using techniques such as autoregressive [26] based on the historical time series $x = \{x_1, \cdots, x_T\}$ to obtain the future time series $\hat{x} = \{x_{T+1}, \cdots, x_{\hat{T}}\}$. The second step is to identify events in the predicted future time series $\hat{x}$ using either unsupervised methods such as burstness detection [31] and change detection [109], or supervised techniques based on learning event characterization function. For example, existing works [165, 173] first represent the predicted future time series $\hat{x} \in \mathbb{R}^{\hat{T} \times D}$ using time-delayed embedding, into $\tilde{x} \in \mathbb{R}^{\hat{T} \times D'}$ where each observation at time t can be represented as $\tilde{x}_t = \{x_{t-(D'-1)\tau}, \cdots, x_{t-2\tau}, x_{t-\tau}, x_t\}$ and $t = T, T+1, \cdots \hat{T}$. Then an event characterization function $f_c(\tilde{x}_t)$ is established to map $\tilde{x}_t$ to the likelihood of an event, which can be fitted based on the event labels provided in the training set intuitively. Overall, the unsupervised method requires users to assume the type of patterns (e.g., burstiness and change) of future events based on prior knowledge but do not require event label data. However, in cases where the event time series

pattern is difficult to assume but the label data is available, supervised learning-based methods are usually used.

3.1.3 Continuous-time Prediction. Discrete-time prediction methods, although usually simple to establish, also suffer from several issues. First, their time-resolution is limited to the discretization granularity; increasing this granularity significantly increases the computations al resources required, which means the resolution cannot be arbitrarily high. Moreover, this trade-off is itself a hyperparameter that is sensitive to the prediction accuracy, rendering it difficult and time-consuming to tune during training. To address these issues, a number of techniques work around it by directly predicting the continuous-valued event time [191], usually by leveraging one of three techniques.
1. Simple Regression. The simplest methods directly formalize continuous-event-time prediction as a regression problem [26], where the output is the numerical-value future event time [212] and/or their duration [80, 129]. Common regressors such as linear regression and recurrent neural networks have been utilized for this. Despite their apparent simplicity, this is not straightforward as simple regression typically assumes Gaussian distribution [129], which does not usually reflect the true distribution of event times. For example, the future event time needs to be left-bounded (i.e., larger than the current time), as well asbeing typically non-symmetric and usually periodic, with recurrent events having multiple peaks in the probability density function along the time dimension.
2. Point Processes. As they allow more flexibility in fitting true event time distributions, point process methods [167, 219] are widely leveraged and have demonstrated their effectiveness for continuous time event prediction tasks. They require a conditional intensity function, defined as follows:

$$\lambda(t|X) = \mathbb{E}[N(t, t + \mathrm{d}t)/\mathrm{d}t|X] = g(t|X)/(1 - G(t|X)) \tag{1}$$

where $g(t|X)$ is the conditional density function of the event occurrence probability at time t given an observation X, and whose corresponding cumulative distribution function, $G(t|X)$), $N(t, t + \mathrm{d}t)$, denotes the count of events during the time period between t and $t + \mathrm{d}t$, where $\mathrm{d}t$ is an infinitely-small time period.

Hence, by leveraging the relation between density and accumulative functions and then rearranging Equation (1), the following conditional density function is obtained:

$$g(t|X) = \lambda(t|X) \cdot \exp \cdot \left(- \int_{t_0}^{t} \lambda(u|X) \mathrm{d}u \right) \tag{2}$$

Once the above model has been trained using a technique such as maximal likelihood [26], the time of the next event in the future is predicted as:

$$\hat{t} = \int_{t_0}^{\infty} t \cdot g(t|X) \mathrm{d}t \tag{3}$$

Although existing methods typically share the same workflow as that shown above, they vary in the way they define the conditional intensity function $\lambda(t|X)$. Traditional models typically utilize prescribed distributions such as the Poisson distribution [191], Gamma distribution [53], Hawks [69], Weibull process [56], and other distributions [219]. For example, Damaschke et al. [56] utilized a Weibull distribution to model volcano eruption events, while Ertekin et al. [72] instead proposed the use of a non-homogeneous Poisson process to fit the conditional intensity function for power system failure events. However, in many other situations where there is no information regarding appropriate prescribed distributions, researchers must start by leveraging nonparametric approaches to learn sophisticated distributions from the data using expressive models such as

neural networks. For example, Simma and Jordan [191] utilized of RNN to learn a highly nonlinear function of $\lambda(t|X)$.

3. Survival Analysis. Survival analysis [62, 204] is related to point processes in that it also defines an event intensity or hazard function, but in this case based on survival probability considerations, as follows:

$$H(t|X) = \left(\xi(t - \mathrm{d}t|X) - \xi(t|X)\right)/\xi(t|X), \text{ where } \xi(t|X) = p(\hat{t} > t) \tag{4}$$

where $H(t|X)$ is the so-called Hazard function denoting the hazard of event occurrence between time $(t - \mathrm{d}t)$ for a t for a given observation X. Either $H(t|X)$ or $\xi(t|X)$ could be utilized for predicting the time of future events. For example, the event occurrence time can be estimated when $\xi(t|X)$ is lower than a specific value. Also, one can obtain $\xi(t|X) = \exp\left(-\int_0^t \mathrm{H(u|X)du}\right)$ according to Equation (4) [132]. Here $H(t|X)$ can adopt any one of several prescribed models, such as the well-known Cox hazard model [61, 132]. To learn the model directly from the data, some researchers have recommended enhancing it using deep neural networks [119]. Vahedian et al. [204] suggest learning the survival probability $\xi(t|X)$ and then applying the function $H(\cdot|X)$ to indicate an event at time t if $H(t|X)$ is larger than a predefined threshold value. A classifier can also be utilized.

Instead of using the raw sequence data, the conditional intensity function can also be projected onto additional continuous-time latent state layers that eventually map to the observations [62, 205]. These latent states can then be extracted using techniques such as hidden semi-Markov models [26], which ensure the elicitation of the continuous time patterns.

3.2 Location Prediction

Event location prediction focuses on predicting the location of future events. Location information can be formulated as one of two types: **1. Raster-based.** Here, a continuous space is partitioned into a grid of cells, each of which represents a spatial region, as shown in Figure 4(a). This type of representation is suitable for situations where the spatial size of the event is non-negligible. **2. Point-based.** In this case, each location is represented by an abstract point with infinitely-small size, as shown in Figure 4(b). This type of representation is most suitable for the situations where the spatial size of the event can be neglected, or the location regions of the events can only be in discrete spaces such as network nodes.

(a) Raster-based event prediction (b) Point-based event prediction in Euclidean and non-Euclidean spaces

Fig. 4. Raster-based and Point-based Event Location Predictions.

3.2.1 Raster-based Location Prediction. There are three types of techniques used for raster-based event location prediction, namely spatial clustering, spatial embedding, and spatial convolution.

1. Spatial clustering. In raster-based representations, each location unit is usually a regular grid cell with the same size and shape. However, regions with similar spatial characteristics typically have irregular shapes and sizes, which could be approximated as composite representations of a number of grids [105]. The purpose of spatial clustering here is to group the contiguous regions

who collectively exhibit significant patterns. The methods are typically agglomerative style. They typically start from the original finest-grained spatial raster units and proceed by merging the spatial neighborhood of a specific unit in each iteration. But different research works define different criteria for instantiating the merging operation. For example, Wang and Ding [211] merge neighborhoods if the unified region after merging can maintain the spatially frequent patterns. Xiong et al. [220] chose an alternative approach by merging spatial neighbor locations into the current locations sequentially until the merged region possesses event data that is sufficiently statistically significant. These methods usually run in a greedy style to ensure their time complexity remains smaller than quadratic. After the spatial clustering is completed, each spatial cluster will be input into the classifier to determine whether or not there is an event corresponding to it.

2. Spatial interpolation. Unlike spatial clustering-based methods, spatial interpolation based methods maintain the original fine granularity of the event location information. The estimation of event occurrence probability can be further interpolated for locations with no historical events and hence achieve spatial smoothness. This can be accomplished using commonly-used methods such as kernel density estimation [5, 93] and spatial Kriging [105, 114]. Kernel density estimation is a popular way to model the geo-statistics in numerous types of events such as crimes [5] and terrorism [93]:

$$k(s) = 1/(n \cdot \gamma) \sum_{i=1}^{n} K\left((s - s_i)/\gamma\right) \tag{5}$$

where $k(s)$ denotes the kernel estimation for the location point s, n is the number of historical event locations, each s_i is a historical event location, γ is a tunable bandwidth parameter, and $K(\cdot)$ is a kernel function such as Gaussian kernel [85].

More recently, Ristea et al. [176] further extended KDE-based techniques by leveraging Localized KDE and then applying spatial interpolation techniques to estimate spatial feature values for the cells in the grid. Since each cell is an area rather than a point, the center of each cell is usually leveraged as the representative of this cell. Finally, a classifier will take this as its input to predict the event occurrence for each grid [5, 176].

3. Spatial convolution. In the last few years, Convolutional Neural Networks (CNNs) have demonstrated significant success in learning and representing sophisticated spatial patterns from image and spatial data [88]. A CNN contains multiple convolutional layers that extract the hierarchical spatial semantics of images. In each convolutional layer, a convolution operation is executed by scanning a feature map with a filter, which results in another smaller feature map with a higher level semantic. Since raster-based spatial data and images share a similar mathematical form, it is natural to leverage CNNs to process it.

Existing methods [19, 150, 164, 209] in this category typically formulate a spatial map as input to predict another spatial map that denotes future event hotspots. Such a formulation is analogous to the "image translation" problem popular in recent years in the computer vision domain [46]. Specifically, researchers typically leverage an encoder-decoder architecture, where the input images (or spatial map) are processed by multiple convolutional layers into a higher-level representation, which is then decoded back into an output image with the same size, through a reverse convolutional operations process known as transposed convolution [88].

4. Trajectory destination prediction. This type of method typically focuses on population-based events whose patterns can be interpreted as the collective behaviors of individuals, such as "gathering events" and "dispersal events". These methods share a unified procedure that typically consists of two steps: 1) predict future locations based on the observed trajectories of individuals, and 2) detect the occurrence of the "future" events based on the future spatial patterns obtained in Step 1. The specific methodologies for each step are as follows:

- **Step 1:** Here, the aim is to predict each location an individual will visit in the future, given a historical sequence of locations visited. This can be formulated as a sequence prediction problem. For example, Wang and Gerber [214] sought to predict the probability of the next time point $t+1$'s location s_{t+1} based on all the preceding time points: $p(s_{t+1}|s_{\leq t}) = p(s_{t+1}|s_t, s_{t-1}, \cdots, s_0)$, based on various strategies including a historical volume-based prior model, Markov models, and multi-class classification models. Vahedian et al. [203] adopted Bayesian theory $p(s_{t+1}|s_{\leq t}) = p(s_{\leq t}|s_{t+1}) \cdot p(s_{t+1})/p(s_{\leq t})$ which requires the conditional probability $p(s_{\leq t}|s_{t+1})$ to be stored. However, in many situations, there is huge number of possible trajectories for each destination. For example, with a 128×64 grid, one needs to store $(128 \times 64)^3 \approx 5.5 \times 10^{11}$ options. To improve the memory efficiency, this can be limited to a consideration of just the source and current locations, leveraging a quad-tree style architecture to store the historical information. To achieve more efficient storage and speed up $p(s_{\leq t}|s_{t+1})$ queries, Vahedian et al. [203] further extended the quad-tree into a new technique called VIGO, which removes duplicate destination locations in different leaves.
- **Step 2:** The aim in this step is to forecast future event locations based on the future visiting patterns predicted in Step 1. The most basic strategy here is to consider each grid cell independently. For example, Wang and Gerber [214] adopted supervised learning strategies to build predictive mapping between the visiting patterns and the event occurrence. A more sophisticated approach is to consider the spatial outbreaks composited by multiple grids. Scalable algorithms have also been proposed to identify regions containing statistically significant hotspots [110], such as spatial scan statistics [116]. Khezerlou et al. [110] proposed a greedy-based heuristic tailored for the grid-based data formulation, which extends the original "seed" grid containing statistically-large future event densities to four directions until the extended region is no longer a statistically-significant outbreak.

3.2.2 Point-based Prediction. . Unlike the raster-based formulation, which covers the prediction of a contiguous spatial region, point-based prediction focuses specifically on locations of interest, which can be distributed sparsely in a Euclidean (e.g., spatial region) or non-Euclidean space (e.g., graph topology). These methods can be categorized into supervised and unsupervised approaches. *1. Supervised approaches.* In supervised methods, each location will be classified as either "positive" or "negative" with regard to a future event occurrence. The simplest setting is based on the independent and identically distributed (i.i.d.) assumption among the locations, where each location is predicted by a classifier independently using their respective input features. However, given that different locations usually have strong spatial heterogeneity and dependency, further research has been proposed to tackle them based on different locations' predictors and outputs, resulting in two research directions: 1) Spatial multi-task learning, and 2) Spatial auto-regressive methods.

- **Spatial multi-task learning.** Multi-task learning is a popular learning strategy that can jointly learn the models for different tasks such that the learned model can not only share their knowledge but also preserve some exclusive characteristics of the individual tasks [244]. This notion coincides very well with spatial event prediction tasks, where combining the outputs of models from different locations needs to consider both their spatial dependency and heterogeneity. Zhao et al. [244] proposed a spatial multi-task learning framework as follows:

$$\min_{W} \sum_{i}^{m} \mathcal{L}(Y_i, f(W_i, X_i)) + \alpha \cdot \mathcal{R}(\{W_i\}_i^k, M), \quad s.t., \ C(\{W_i\}_i^m, D) \in \mathbb{C} \tag{6}$$

where m is the total number of locations (i.e., tasks), W_i and Y_i are the model parameters and true labels (event occurrence for all time points), respectively, of task i. $\mathcal{L}(\cdot)$ is the empirical loss, $f(W_i, X_i)$ is the predictor for task i, and $\mathcal{R}(\cdot)$ is the spatial regularization term

based on the spatial dependency information $M \in \mathbb{R}^{m \times m}$, where $M_{i,j}$ records the spatial dependency between location i and j. $C(\cdot)$ represents the spatial constraints imposed over the corresponding models to enforce them to remain within the valid space $\mathbb{C}$. Over recent years, there have been multiple studies proposing different strategies for $\mathcal{R}(\cdot)$ and $C(\cdot)$. For example, Zhao et al. [245] assumed that all the locations would be evenly correlated and enforced their similar sparsity patterns for feature selection, while Gao et al. [85] further extended this to differentiate the strength of the correlation between different locations' tasks according to the spatial distance between them. This research has been further extended this approach to tree-structured multitask learning to handle the hierarchical relationship among locations at different administrative levels (e.g., cities, states, and countries) [246] in a model that also considers the logical constraints over the predictions from different locations who have hierachical relationships. Instead of evenly similar, Zhao, et al. [243] further estimated spatial dependency D utilizing inverse distance using Gaussian kernels, while Ning et al. [156] proposed estimating the spatial dependency D based on the event co-occurrence frequency between each pair of locations.

- **Spatial auto-regressive methods**. Spatial auto-regressive models have been extensively explored in domains such as geography and econometrics, where they are applied to perform predictions where the i.i.d. assumption is violated due to the strong dependencies among neighbor locations. Its generic framework is as follows:

$$\hat{Y}_{t+1} = \rho M \hat{Y}_{t+1} + X_t \cdot W + \varepsilon \tag{7}$$

where $X_t \in \mathbb{R}^{m \times D}$ and $\hat{Y}_{t+1} \in \mathbb{R}^{m \times m}$ are the observations at time t and event predictions at time $t+1$ over all the m locations, and $M \in \mathbb{R}^{m \times m}$ is the spatial dependency matrix with zero-valued diagonals. This means the prediction of each location $\hat{Y}_{t+1,i} \in \hat{Y}_{t+1}$ is jointly determined by its input $X_{t,i}$ and neighbors $\{j | M_{i,j} \neq 0\}$ and ρ is a positive value to balance these two factors. Since event occurrence requires discrete predictions, simple threshold-based strategies can be used to discretize $\hat{Y}_i$ into $\hat{Y}'_i = \{0, 1\}$ [32]. Moreover, due to the complexity of event prediction tasks and the large number of locations, sometimes it is difficult to define the whole M manually. Zhao et al. [243] proposed jointly learning the prediction model and spatial dependency from the data using graphical LASSO techniques. Yi et al. [228] took a different approach, leveraging conditional random fields to instantiate the spatial autoregression, where the spatial dependency is measured by Gaussian kernel-based metrics. Yi et al. [227] then went on to propose leveraging the neural network model to learn the locations' dependency.

2. Unsupervised approaches. Without supervision from labels, unsupervised-based methods must first identify potential precursors and determinant features in different locations. They can then detect anomalies that are characterized by specific feature selection and location combinatorial patterns (e.g., spatial outbreaks and connected subgraphs) as the future event indicators [41]. The generic formulation is as follows:

$$(F, R) = \arg\max_{F,R} \; q(F, R) \quad s.t., \; \text{supp}(F_i) \in \mathbb{M}(\mathbb{G}, \beta), \; \text{supp}(R_i) \in \mathbb{C} \tag{8}$$

where $q(\cdot)$ denotes scan statistics which score the significance of each candidate pattern, represented by both a candidate location combinatorial pattern R and feature selection pattern F. Specifically, $F \in \{0, 1\}^{D' \times n}$ denotes the feature selection results (where "1" means selected; "0", otherwise) and $R \in \{0, 1\}^{m \times n}$ denotes the m involved locations for the n events. $\mathbb{M}(\mathbb{G}, \beta)$ and $\mathbb{C}$ are the set of all the feasible solutions of F and R, respectively. $q(\cdot)$ can be instantiated by scan statistics such as Kulldorff's scan statistics [116] and the Berk-Jones statistic [41], which can be applied to detect and forecast events such as epidemic outbreaks and civil unrest events [171]. Depending on whether the

embedding space is an Euclidean region (e.g., a geographical region) or a non-Euclidean region (e.g., a network topology), the pattern constraint C can be either constrained to predefined geometric shapes such as a circle, rectangle, or an irregular shape or subgraphs such as connected, cliques, and k-cliques. The problem in Equation (8) is nonconvex and sometimes even discrete, and hence difficult to solve. A generic way is to optimize F using sparse feature selection; there is a useful survey provided in [127] and R can be defined using the two-step graph-structured matching method detailed in[42]. More recently, new techniques have been developed that are capable of jointly learning both feature and location selection [42, 187].

3.3 Semantic Prediction

Event semantics prediction addresses the problem of forecasting topics, descriptions, or other meta-attributes in addition to future events' times and locations. Unlike time and location prediction, the data in event semantics prediction usually involves symbols and natural languages in addition to numerical quantities, which means different types of techniques may be utilized. The data are categorized into three types based on how the historical data are organized and utilized to infer future events. The first of these categories covers rule-based methods, where future event precursors are extracted by mining association or logical patterns in historical data. The second type is sequence-based, considering event occurrence to be a consequence of temporal event chains. The third type further generalizes event chains into event graphs, where additional cross-chain contexts need to be modeled. These are discussed in turn below.

3.3.1 Association-based Prediction. . Association rule-based methods are amongst the most classic approaches in data mining domain for event prediction, typically consisting of two steps: 1) learn the associations between precursors and target events, and then 2) utilize the learned associations to predict future events. For the first step, for example, an association could be x = {"election", "fraud"} $\rightarrow$ y ="protest event", which indicates that serious fraud occurring in an election process could lead to future protest events. To discover all the significant associations from the ocean of candidate rules efficiently, frequent set mining [92] can be leveraged. Each discovered rule needs to come with both sufficient *support* and *confidence*. Here, *support* is defined as the number of cases where both "x" and "y" co-occur, while *confidence* means the ratio indicating that "y" occurs once "x" happens. To better estimate these discrimination rules, further temporal constraints can be added that require the occurrence time of "x" and "y" to be sufficiently close to be considered "co-occurrences". Once the frequent set rules have been discovered, pruning strategies may be applied to retain the most accurate and specific ones, with various strategies for generating final predictions [92]. Specifically, given each new observation x', one of the simplest strategies is to output the events that are triggered by any of the association rules starting from event x' [206]. Other strategies first rank the predicted results based on their confidence and then predict just the top r events [252]. More sophisticated and rigorous strategies tend to build a decision list where each element in the list is an association rule mapping, so once a generative model has been built for the decision process, the maximal likelihood can be leveraged to optimize the order of the decision list [124].

3.3.2 Causality-based prediction. This type of research leverages the causality inferred among the historical events to achieve future event predictions. The data here typically shares a generic framework consisting of the following procedures: 1) event representation, 2) event graph construction, and 3) future event inference.

Step 1: Event semantic representation. This approach typically begins by extracting the events from the target texts using natural language processing techniques such as sanitization, tokenization, POS tag analysis, and name entity recognition. Several types of objects can be extracted to represent

the events: i) Noun Phrase-based [39, 94, 111], where the noun-phrase corresponds to each event (for example, "2008 Sichuan Earthquake"); ii) Verbs and Nouns [112, 168], where an event is represented as a set of noun-verb pairs extracted from news headlines (for example, "<capture, people>", "<escape, prison>", or "<send, prison>"); and iii) Tuple-based [249], where each event is represented by a tuple consisting of objects (such as actors, instruments, or receptors), a relationship (or property), and time. An RDF-based format has also been leveraged in some works [57].

Step 2: Event causality inference. The goal here is to infer the cause-effect pairs among historical events. Due to its combinatorial nature, narrowing down the number of candidate pairs is crucial. Existing works usually begin by clustering the events into event chains, each of which consist of a sequence of time-ordered events under the relevant semantics, typically the same topics, actors, and/or objects [2]. The causal relations among the event pairs can then be inferred in various ways. The simplest approach is just to consider the likelihood that y occurs after x has occurred throughout the training data. Other methods utilize NLP techniques to identify causal mentions such as causal connectives, prepositions, and verbs [168]. Some formulate causal-effect relationship identification as a classification task where the inputs are the cause and effect candidate events, often incorporating contextual information including related background knowledge from web texts. Here, the classifier is built on a multi-column CNN that outputs either "1" or "0" to indicate whether the candidate has an effect or not [115]. In many situations, the cause-effect rules learned directly using the above methods can be too specific and sparse, with low generalizability, so a typical next step is to generalize the learned rules. For example, "Earthquake hits China" → "Red Cross help sent to Beijing" is a specific rule that can be generalized to "Earthquake hits [A country]" → "Red Cross help sent to [The capital of this country]". To achieve this, some external ontology or a knowledge base is typically needed in order to establish the underlying relationships among items or provide necessary information on their properties, such as Wikipedia (https://www.wikipedia.org/), YAGO [196], WordNet [75], or ConceptNet [137]. Based on these resources, the similarity between two cause-effect pairs (c_i, ε_i) and (c_j, ε_j) can be computed by jointly considering the respective similarity of the putative cause and effect: $\sigma((c_i, \varepsilon_i), (c_j, \varepsilon_j)) = (\sigma(c_i, c_j) + \sigma(\varepsilon_i, \varepsilon_j))/2$. An appropriate algorithm can then be utilized to apply hierarchical agglomerative clustering to group them and hence generate a data structure that can efficiently manage the task of storing and querying them to identify any cause-effect pairs. For example, [168, 169, 190] leverage an abstraction tree, where each leaf is an original specific cause-effect pair and each intermediate node is the centroid of a cluster. Instead of using hierarchical clustering, [249] directly uses the word ontology to simultaneously generalize cause and effect (e.g., the noun "violet" is generalized to "purple", the verb "kill" is generalized to "murder-42.1[1]") and then leverage a hierarchical causal network to organize the generalized rules.

Step 3: Future event Inference. Given an arbitrary query event, two steps are needed to infer the future events caused by it based on the causality of events learned above. First, we need to retrieve similar events that match the query event from historical event pool. This requires the similarity between the query event and all the historical events to be calculated. To achieve this, Lei et al. [123] utilized context information, including event time, location, and other environmental and descriptive information. For methods requiring event generalization, the first step is to traverse the abstraction tree starting from the root that corresponds to the most general event rule. The search frontier then moves across the tree if the child node is more similar, culminating in the nodes which are the least general but still similar to the new event being retrieved [168]. Similarly, [45] proposed another tree structure referred to as a "circular binary search tree" to manage the event occurrence pattern. We can now apply the learned predicate rules starting from the retrieved event

[1]the form of verb class in VerbNet [249].

to obtain the prediction results. Since each cause event can lead to multiple events, a convenient way to determine the final prediction is to calculate the support [168], or conditional probability [226] of the rules. Radinsky et al. [168] took a different approach, instead ranking the potential future events by their similarity defined by the length of their minimal generalization path. For example, the minimal generalization path for "London" and "Paris" is "London" $\xrightarrow{\text{capital-of}}$ "Great Brain" $\xrightarrow{\text{in-continent}}$ "Europe" $\xleftarrow{\text{in-continent}}$ "France" $\xleftarrow{\text{capital-of}}$ "Paris". Alternatively, Zhao et al. [249] proposed embedding the event causality network into a continuous vector space and then applying an energy function designed to rank potential events, where true cause-effect pairs are assumed to have low energies.

3.3.3 Semantic Sequence. These methods share a very straightforward problem formulation. Given a temporal sequence for a historical event chain, the goal is to predict the semantics of the next event using sequence prediction [26]. The existing methods can be classified into four major categories: 1) classical sequence prediction; 2) recurrent neural networks; 3) Markov chains; and 4) time series predictions.

Sequence classification-based methods. These methods formulate event semantic prediction as a multi-class classification problem, where a finite number of candidate events are ranked and the top-ranked event is treated as the future event semantic. The objective is $\hat{C} = \arg\max_{C_i} u(s_{T+1} = C_i|s_1, \cdots, s_T)$, where s_{T+1} denotes the event semantic in time slot $T+1$ and $\hat{C}$ is the optimal semantic among all the semantic candidates C_i $(i = 1, \cdots)$. Multi-class classification problems can be split into events with different topics/semantic meaning. Three types of sequence classification methods have been utilized for this purpose, namely feature-based methods, prototype-based methods, and model-based methods such as Markov models.

- **Feature-based.** One of the simplest methods is to ignore the temporal relationships among the events in the chain, by either aggregating the inputs or the outputs. Tama and Comuzzi [198] formulated historical event sequences with multiple attributes for event prediction, testing multiple conventional classifiers. Another type of approach based on this notion utilizes compositional based-methods [89] that typically leverage the assumption of independency among the historical input events to simplify the original problem $u(s_{T+1}|s_1, s_2, \cdots, s_T) = u(s_{T+1}|s_{\leq T})$ into $v(u(s_{T+1}|s_1), u(s_{T+1}|s_2), \cdots, u(s_{T+1}|s_T))$ where $v(\cdot)$ is simply an aggregation function that represents a summation operation over all the components. Each component function $u(s_{T+1}|s_i)$ can then be calculated by estimating how likely it is that event semantic s_{T+1} and s_i $(i \leq T)$ co-occur in the same event chain. Granroth-Wilding and Clark [89] investigated various models ranging from straightforward similarity scoring functions through bigram models and word embedding combined with similarity scoring functions to newly developed composition neural networks that jointly learn the representation of s_{T+1} and s_i and then calculate their coherence. Some other researchers have gone further to consider the dependency among the historical events. For example, Letham et al. [125] proposed to optimizing the correct ordering among the candidate events, based on the following equation:

$$\sum_{i\in\mathcal{I},\ j\in\mathcal{J}} \mathbb{1}_{[u(s_{T+1}=C_i|s_{\leq T})>u(s_{T+1}=C_j|s_{\leq T})]} \implies \sum_{i\in\mathcal{I},\ j\in\mathcal{J}} e^{u(s_{T+1}=C_i|s_{\leq T})-u(s_{T+1}=C_j|s_{\leq T})} + \rho\|W\|_2^2$$

where the semantic candidate in the set $\mathcal{I}$ should be ranked strictly to be lower than those in $\mathcal{J}$, with the goal being to penalize the "incorrect ordering". Here, $\mathbb{1}_{[\cdot]}$ is an indicator function which is discrete such that $\mathbb{1}_{[b\geq a]} \leq e^{b-a}$ and can thus be utilized as the upper-bound for minimization, as can be seen in the right-hand-side of the above equation. W is the set of parameters of the function $u(\cdot)$. This can now be relaxed to an exponential-based approximation for effective optimization using gradient-based algorithms [88]. Other

methods focus on first transferring the sequential data into sequence embeddings that can encode the latent sequential context. For example, Fronza et al. [79] apply random indexing to represent the words in terms of their its vector representations by embedding the information from neighboring words into each word before utilizing conventional classifiers such as Support Vector Machines (SVM) to identify the future events.

- **Model-based.** Markov-based models have also been leveraged to characterize temporal patterns [224]. These typically use E_i to denote each event under a specific type and $\mathcal{E}$ denotes the set of event types. The goal here is to predict the event type of the next event to occur in the future. In [7], the event types are modeled using the Markov model so given the current event type, the next event type can be inferred simply by looking up the state with the highest probability in the transition matrix. A tool called Wayeb [8] has been developed based on this method. Laxman et al. [121] developed a more complicated model, based on a mixture of Hidden Markov models and introducing new assumptions and the concept of episodes composed of a subsequence of event types. They assumed different event episodes should have different transition patterns so started by discovering the frequent episodes for events, each of which they modeled by a specific hidden Markov model over various event types. This made it possible to establish the generative process for each future event type s based on the mixture of the above episode Markov models. When predicting, the likelihood of a current observed event sequence over each possible generative process, $p(X|\Lambda_Y)$ is evaluated, after which a future event type can be considered as either being larger than some threshold (as in [121]) or the largest among all the different Y values (in [239, 241]).
- **Prototype-based.** Adhikari et al. [3] took a different approach, utilizing a prototype-based strategy that first clusters the event sequences into different clusters in terms of their temporal patterns. When a new event sequence is observed, its closest cluster's centroid will then be leveraged as a "reference event sequence" whose sub-sequential events will be referred to when predicting future events for this new event sequence.

Recurrent neural network (RNN)-based methods. Approaches in this category can be classified into two types: 1. Attribute-based models; and 2. Descriptive-based models. The attribute-based models, ingest feature representation of events as input, while the descriptive-based models typically ingest unstructured information such as texts to directly predict future events.

- **Attributed-based Methods.** Here, each event $y = (t, l, s)$ at time t is recast and represented as $e_t = (e_{t,1}, e_{t,2}, \cdots, e_{t,k})$, where $e_{t,i}$ is the i-th feature of the event at time t. The feature here can include location and other information such as event topic and semantics. Each sequence $e = (e_1, \cdots, e_t)$ is then input into the standard RNN architecture for predicting next event e_{t+1} in the sequence at time point $t + 1$ [134]. Various types of RNN components and architecture have been utilized for this purpose [33, 34], but a vanilla RNN [70, 88] for sequence-based event prediction can be written in the following form:

$$h_i = \tanh(a_t),\ a_i = b + W \cdot h_{i-1} + U \cdot e_i,\ o_i = c + V \cdot h_i,\ \psi(i+1) = \text{softmax}(o_i),\ i \leq t$$

where h_i, o_i, and a_i are the latent state, output, and activation for the i-th event, respectively, and W, U, and V are the model parameters for fitting the corresponding mappings. The prediction $e_{t+1} := \psi(t + 1)$ can then be calculated in a feedforward way from the first event and the model training can be done by back-propagating the error from the layer of $\psi(t)$. Existing work typically utilizes the variants of vanilla RNN to handle the gradient vanishing problem, especially when the event chain is not short. The most commonly used methods for event prediction are LSTM and GRU [88]. For example, the architecture and equation for

LSTM are as follows:

$$a_i = \sigma(W_j \cdot [h_{i-1}, e_i] + b_j),\ \tilde{C}_i = \tanh(W_C \cdot [h_{i-1}, h_i] + b_C),\ C_i = \zeta_i C_{i-1} + a_i \tilde{C}_i, \\ \zeta_i = \sigma(W_\zeta \cdot [h_{i-1}, e_i] + b_\zeta),\ o_i = \sigma(W_o[h_{i-1}], e_i) + b_o),\ h_i = o_i * \tanh(C_i) \tag{9}$$

where the additional components C_{i-1} and ζ_i are introduced to keep tracking the previous "history" and gating the information for forgetting in order to handle longer sequences. For example, some researchers opt to leverage a simple type LSTM architecture to extend the RNN-based sequential event prediction [33, 97], while others leverage variants of LSTM, such as bi-directional LSTM instead [113, 155]and yet others prefer to leverage gated-recurrent units (GRU) [70].

Moving beyond considering just the chain relationships among events, Li et al. [131] generalized this into graph-structured relationships to better incorporate the event contextual information via the Narrative Event Evolutionary Graph (NEEG). An NEEG is a knowledge graph where each node is an event and each edge denotes the association between a pair of events, enabling the NEEG to be represented by a weighted adjacency matrix A. The basic architecture can be denoted by the following, as detailed in the paper [131]:

$$a_i = A^{\mathrm{T}} h_{i-1} + b,\ z_i = \sigma(W_z a_i + U_z h_{i-1}),\ r_i = \sigma(W_r a_i + U_r h_{i-1}), \\ c_i = \tanh(W a_i + U(r_i h_{i-1})),\ h_i = (1 - z_i) h_{i-1} + z_i c_i \tag{10}$$

Here, the current activation a_i is not only dependent on the previous time point but also influenced by its neighbor nodes in NEEG.

- **Descriptive-based Methods.** Attribute-based methods require extra effort during preprocessing in order to convert the unstructured raw data into feature vectors, a process which is not only computationally labor intensive but also not always feasible. Therefore, multiple architectures have been proposed to directly process the raw (textual) event descriptions to enable them to be used to predict future event semantics or descriptions. These models share a similar generic framework [96, 97, 139, 195, 221, 231], which begins by encoding each sequence of words into event representations, utilizing an RNN architecture, as shown in Figure 5. The sequence of events must then be characterized by another higher-level RNN to predict future events. Under this framework, some works begin by decoding the predicted future candidate events into event embedding, after which they are compared with each other and the one with the largest confidence score is selected as the predicted event. These methods are usually constrained by the known list of event types, but sometimes we are interested in open set predictions where the predicted event type can be a new appearance of a type that has not previously been seen in the training set. To achieve this, other methods focus on directly generating future events' descriptions that characterize event semantics that may or may not have appeared before by designing an additional sequence decoder that decodes the latent representation of future events into word sequences. More recent research has enhanced the utility and interpretability of the relationship between words and relevant events, and all the previous events for the relevant future event, by adding a hierarchical attention mechanisms. For example, Yu et al. [231] and Su and Jiang [195] both proposed word-level attention and event-level attention, while Hu [96] leveraged word-level attention in the event encoder as well as in the event decoder.

3.4 Multi-faceted Prediction

This section discusses the research into ways to jointly predict the time, location, and semantics of future events. Existing work in this area can be categorized into three types: 1) joint time and

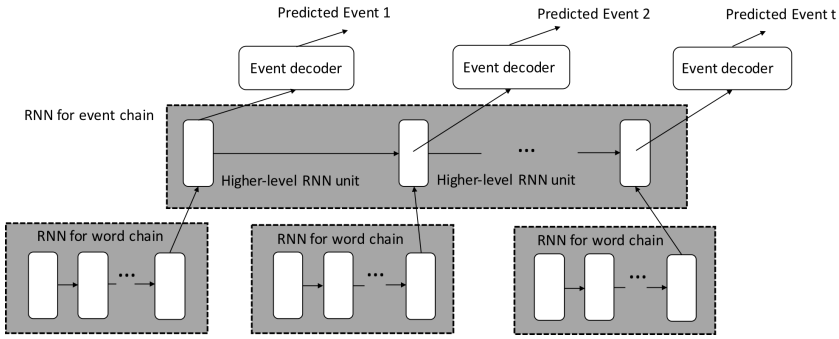

Fig. 5. Generic framework for hierarchical RNN-based event forecasting.

semantics prediction; 2) joint time and location prediction; and 3) joint time, location, and semantic prediction.

3.4.1 Time and Semantics. . For joint time and semantic prediction, there are three popular types of methods, discussed in turn below

Temporal association rule. A temporal association rule can be developed from the vanilla association rule $LHS \rightarrow RHS$ by embedding additional temporal information into either LHS, RHS or both, thus redefining the meaning of co-occurrence and association with temporal constraints. For example, Vilalta and Ma [206] defined LHS as a tuple (E_L, τ), where τ is the time window before the target in RHS predefined by the user. Only the events occurring within a time window before the event in RHS will satisfy the LHS. Similar techniques have also been leveraged by other researchers [45, 194]. However, τ is difficult to define beforehand and it is preferable to be flexible to suit different target events. To handle this challenge, Yang et al. [225] proposed a way to automatically identify information on a continuous time interval from the data. Here, each transaction is composed of not only items but also continuous time duration information. LHS is a set of items (e.g., previous events) while RHS is a tuple $(E_R, [t_1, t_2])$ consisting of a future event semantic representation and its time interval of occurrence. To automatically learn the time interval in RHS, [225] proposed the use of two different methods . The first is called the *confidence-interval-based* method, which leverages a statistical distribution (e.g., Gaussian and student-t [26]) to fit all the observed occurrence times of events in RHS, and then treats the statistical confidence interval as the time interval. The second method is known as *minimal temporal region selection*, which aims to find the temporal region with the smallest interval and covers all historical occurrences of the event in RHS.

Time expression extraction. In contrast to the above statistical-based methods, another way to achieve event time and semantics joint prediction comes from the pattern recognition domain, aiming to directly discover time expressions that mention the (planned) future events. As this type of technique can simultaneously identify time, semantics, and other information such as locations, it is widely used and will be discussed in more details later as part of the discussion of "Planned future event detection methods" in Section 3.4.3.

Time series forecasting-based methods. The methods based on time series forecasting can be separated into direct methods and indirect methods. *Direct methods* typically formulate the event semantic prediction problem as a multi-variate time series forecasting problem, where each variable corresponds to an event type C_i $(i = 1, \cdots)$ and hence the predicted event type at future time $\hat{t}$ is calculated as $\hat{s}_{\hat{t}} = \arg\max_{C_i} f(s_{\hat{t}} = C_i|X)$. For example, in [128], a longitudinal support vector regressor is utilized to predict multi-attribute events, where n support vector regressors, each of which corresponds to an attribute, is built to achieve the goal of predicting the next time point's

attribute value. Weiss and Page [219] took a different approach, leveraging multiple point process models to predict multiple event types. To further estimate the confidence of their predictions, Biloš et al. [25] first leveraged RNN to learn the historical event representation and then input the result into a Gaussian process model to predict future event types. To better capture the joint dynamics across the multiple variables in the time series, Brandt et al. [30] extended this to Bayesian vector autoregression. Utilizing *indirect-style* methods, they focused on learning a mapping from the observed event semantics down to low-dimensional latent-topic space using tensor decomposition-based techniques. Similarly, Matsubara et al. [142] proposed a 3-way topic analysis of the original observed event tensor $Y_0 \in \mathbb{R}^{D_o \times D_a \times D_c}$ consisting of three factors, namely actors, objects, and time. They then went on to decompose this tensor into latent variables via three corresponding low-rank matrices $P_o \in \mathbb{R}^{D_k \times D_o}$, $P_a \in \mathbb{R}^{D_k \times D_a}$, and $P_c \in \mathbb{R}^{D_k \times D_c}$ respectively, as shown in Figure 6. Here D_k is the number of latent topics. For the prediction, the time matrices P_c are predicted into the future $\hat{P}_c$ via multi-variate time series forecasting, after which a future event tensor are estimated by recovering a "future event tensor" $\hat{Y}$ by the multiplication among the predicted time matrix $\hat{P}_c$ as well as the known actor matrix P_a and object matrix P_o.

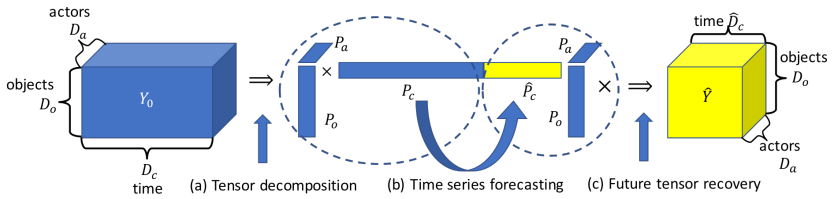

Fig. 6. Tensor Decomposition and Forecasting for Complex Time-stamped events.

3.4.2 Time and Location Prediction. This category of methods focuses on jointly predicting the location and time of future events. These methods can be classified into two subtypes: 1) raster-based: which focus on predictions for individual time slots and location regions, and 2) point-based: which predicts continuous time and location points.

Raster-based. These methods usually formulate data into temporal sequences consisting of spatial snapshots. Over the last few years, various techniques have been proposed to characterize the spatial and temporal information for event prediction.

The simplest way to consider spatial information is to directly treat location information as one of the input features, and then feed it into predictive models, such as linear regression [250], LSTM [174] and Gaussian processes [118]. During model training, Zhao and Tang [250] leveraged the spatiotemporal dependency to regularize their model parameters. Most of the methods in this domain aim to jointly consider the spatial and temporal dependency for predictions [64]. At present, the most popular framework is the CNN+RNN architecture, which implements sequence-to-sequence learning problems such as the one illustrated in Figure 7. Here, the multi-attributed spatial information for each time point can be organized as a series of multi-channel images, which can be encoded using convolution-based operations. For example, Huang et al. [99] proposed the addition of convolutional layers to process the input into vector representations. Other researchers have leveraged variational autoencoders [215] and CNN autoencoders [104] to learn the low-dimensional embedding of the raw spatial input data. This allows the learned representation of the input to be input into the temporal sequence learning architecture. Different recurring units have been investigated, including RNN, LSTM, convLSTM, and stacked-convLSTM [88]. The resulting representation of the input sequence is then sent to the output sequence as input. Here, another recurrent architecture is established. The output of the unit for each time point will be input into a

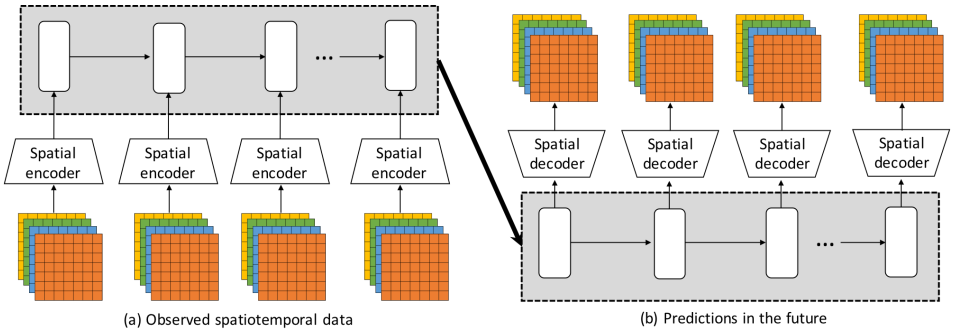

Fig. 7. Generic framework for spatiotemporal event prediction using CNN+RNN-based deep learning framework.

spatial decoder component which can be implemented using transposed convolutional layers [233], transposed convLSTM [104], or a spatial decoder in a variational autoencoder [215]. A conditional random field is another popular technique often used to model the spatial dependency [105].

Point-based. The spatiotemporal point process is an important technique for spatiotemporal event prediction as it models the rate of event occurrence in terms of both spatial and time points. It is defined as:

$$\lambda(t, l|X) = \lim_{|\mathrm{d}t|\to 0, |\mathrm{d}l|\to 0} \mathbb{E}[N(\mathrm{d}t \times \mathrm{d}l)|X]/(|\mathrm{d}t||\mathrm{d}l|) \tag{11}$$

Various models have been proposed to instantiate the model of the framework illustrated in Equation (11). For example, Liu and Brown et al. [136] began by assuming there to be a conditional independence among spatial and temporal factors and hence achieved the following decomposition:

$$\lambda(t, l|X) = \lambda(t, l|L, T, F) = \lambda_1(l|L, T, F, t) \cdot \lambda_2(t|T) \tag{12}$$

where X, L, T, and F denotes the whole input indicator data as well as its different facets, including location, time, and other semantic features, respectively. Then the term $\lambda_1(\cdot)$ can be modeled based on the Markov spatial point process while $\lambda_2(\cdot)$ can be characterized using temporal autoregressive models. To handle situations where explicit assumptions for model distributions are difficult, several methods have been proposed to involve the deep architecture during the point process. Most recently, Okawa et al. [159] have proposed the following:

$$\lambda(t, l|X) = \int g_\theta\left(t', l', \mathcal{F}(t', l')\right) \cdot \mathcal{K}((t, l), (t', l'))\, \mathrm{d}t' \mathrm{d}l' \tag{13}$$

where $\mathcal{K}(\cdot, \cdot)$ is a kernel function such as a Gaussian kernel [26] that measures the similarity in time and location dimensions. $\mathcal{F}(t', l') \subseteq F$ denotes the feature values (e.g., event semantics) for the data at location l' and time t'. $g_\theta(\cdot)$ can be a deep neural network that is parameterized by θ and returns an nonnegative scalar. The model selection of $g_\theta(\cdot)$ depends on the specific data types. For example, these authors constructed an image attention network by combining a CNN with the spatial attention model proposed by Lu et al. [138].

3.4.3 Time, location, and Semantics. In this section, we introduce the strategies that jointly predict the time, location, and semantics of future events, which can be grouped into either system-based or model-based strategies.

System-based. The first type of the system-based methods considered here is the model-fusion system. The most intuitive approach is to leverage and integrate the aforementioned techniques for time, location, and semantics prediction into an event prediction system. For example, a system named EMBERS [171] is an online warning system for future events that can jointly predict the

time, location, and semantics including the type and population of future events. This system also provides information on the confidence of the predictions obtained. Using an ensemble of predictive models for time [160], location, and semantic prediction, this system achieves a significant performance boost in terms of both precision and recall. The trick here is to first prioritize the precision of each individual prediction model by suppressing their recall. Then, due to the diversity and complementary nature of the different models, the fusion of the predictions from different models will eventually result in a high recall. A Bayesian fusion-based strategy has also been investigated [95]. Another system named *Carbon* [108] also leverages a similar strategy.

The second type of model involves crowd-sourced systems that implement fusion strategies to generate the event predictions made by the human predictors. For example, in order to handle the heterogeneity and diversity of the human predictors' skill sets and background knowledge under limited human resources, Rostami et al. [177] proposed a recommender system for matching event forecasting tasks to human predictors with suitable skills in order to maximize the accuracy of their fused predictions. Li et al. [126] took a different approach, designing a prediction market system that operates like a futures market, integrating information from different human predictors to forecast future events. In this system, the predictors can decide whether to buy or sell the "tokens" (using virtual dollars, for example) for each specific prediction they have made according to their confidence in it. They typically make careful decisions as they will obtain corresponding awards (for correct predictions) or penalties (for erroneous predictions).

Planned future event detection methods. These methods focus on detecting the planned future events, usually from various media such sources as social media and news and typically relying on NLP techniques and linguistic principles. Existing methods typically follow a workflow similar to the one shown in Figure 8, consisting of four main steps: 1) *Content filtering*. Methods for content filtering are typically leveraged to retain only the texts that are relevant to the topic of interest. Existing works utilize either supervised methods (e.g., textual classifiers [117] or unsupervised methods (e.g., querying techniques [152, 238]); 2) *Time expression identification* is then utilized to identify future reference expressions and determine the *time to event*. These methods either leverage existing tools such as the Rosetta text analyzer [55] or propose dedicated strategies based on linguistic rules [101]; 3) *Future reference sentence extraction* is the core of planned event detection, and is implemented either by designing regular expression-based rules [153] or by textual classification [117]; and 4) *Location identification*. The expression of locations is typically highly heterogeneous and noisy. Existing works have relied heavily on geocoding techniques that can resolve the event location accurately. In order to infer the event locations, various types of locations are considered by different researchers, such as article locations [152], authors' profile locations [49], locations mentioned in the articles [22], and authors' neighbors' locations [107]. Multiple locations have been selected using a geometric median [49] or fused using logical rules such as probabilistic soft logic [152].

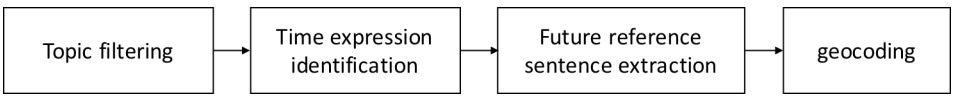

Fig. 8. Generic framework for planned event detection.

Tensor-based methods. Some methods formulate the data into tensor-form, with dimensions including location, time, and semantics. Tensor decomposition is then applied to approximate the original tensors as the product of multiple low-rank matrices, each of which is a mapping from latent topics to each dimension. Finally, the tensor is extrapolated towards future time periods by various strategies. For example, Mirtaheri [148] extrapolated the time dimension-matrix only, which they then multiplied with the other dimensions' matrices to recover the estimated extrapolated tensor

into the future. Zhou et al. [253] took a different approach, choosing instead to add "empty values" for the entries in future time to the original tensor, and then use tensor completion techniques to infer the missing values corresponding to future events.

4 APPLICATIONS OF EVENT PREDICTIONS

4.1 Healthcare

This category generally consists of two types of event prediction: 1) population level, which includes disease epidemics and outbreaks, and 2) individual level, which relates to clinical longitudinal events.

4.1.1 Population-level. There has been extensive research on disease outbreaks for many different types of diseases and epidemics, including seasonal flu [3], Zika [200], H1N1 [158], Ebola [13], and COVID-19 [162]. These predictions target both the location and time of future events, while the disease type is usually fixed to a specific type for each model. Compartmental models such as SIR models are among the classical mathematical tools used to analyze, model, and simulate the epidemic dynamics [186, 237]. More recently, individual-based computational models have begun to be used to perform network-based epidemiology based on network science and graph-theoretical models, where an epidemic is modeled as a stochastic propagation over an explicit interaction network among people [52]. Thanks to the availability of high-performance computing resources, another option is to construct a "digital twin" of the real world, by considering a realistic representation of a population, including membersâĂŹ demographic, geographic, behavioral, and social contextual information, and then using individual-based simulations to study the spread of epidemics within each network [27]. The above techniques heavily rely on the model assumptions regarding how the disease progresses individually and is transmitted from person to person [27]. The rapid growth of large surveillance data and social media data sets such as Twitter and Google flu trends in recent years has led to a massive increase of interest in using data-driven approaches to directly learn the predictive mapping [3]. These methods are usually both more time-efficient and less dependent on assumptions, while the aforementioned computational models are more powerful for longer-term prediction due to their ability to take into account the specific disease mechanisms [242]. Finally, there have also been reports of synergistic research that combines both techniques to benefit from their complementary strengths [98, 242].

4.1.2 Individual-level. This research thread focuses on the longitudinal predictive analysis of individual health-related events, including death occurrence [62], adverse drug events [185], sudden illnesses such as strokes [124] and cardiovascular events [23], as well as other clinical events [62] and life events [59] for different groups of people, including elders and people with mental disease. The goal here is usually to predict the time before an event occurs, although some researchers have attempted to predict the type of event. The data sources are essentially the electronic health records of individual patients [172, 185]. Recently, social media, forum, and mobile data has also been utilized for predicting drug adverse events [185] and events that arise during chronic disease (e.g., chemical radiation and surgery) [62].

4.2 Media

This category focuses on predicting events based on information held in various types of media including: video-based, audio-based, and text-based formats. The core issue is to retrieve key information related to future events utilizing semantic pattern recognition from the data.

4.2.1 Video- and Audio- based. . While event detection has been extensively researched for video data [129] and audio mining [192], event prediction is more challenging and has been attracting

increasing attention in recent years. The goal here is usually to predict the future status of the objects in the video, such as the next action of soccer players [60] or basketball players [154], or the movement of vehicles [235].

4.2.2 Text- and script- based. A huge amount of news data has accumulated in recent decades, much of which can be used for big data predictive analytics among news events. A number of researchers have focused on predicting the location, time, and semantics of various events. To achieve this, they usually leverage the immense historical news and knowledge base in order to learn the association and causality among events, which is then applied to forecast events when given current events. Some studies have even directly generated textual descriptions of future events by leveraging NLP techniques such as sequence to sequence learning [57, 97, 123, 153, 168, 170, 190, 195, 221].

4.3 Transportation

. This category can be classified into: 1) population based events, including dispersal events, gathering events, and congestion; and 2) individual-level events, which focus on fine-grained patterns such as human mobility behavior prediction.

4.3.1 Group transportation pattern. . Here, researchers typically focus on transportation events such as congestion [43, 104], large gatherings [203], and dispersal events [204]. The goal is thus to forecast the future time period [80] and location [203] of such events. Data from traffic meters, GPS, and mobile devices are usually used to sense real-time human mobility patterns. Transportation and geographical theories are usually considered to determine the spatial and temporal dependencies for predicting these events.

4.3.2 Individual transportation behavior. Another research thread focuses on individual-level prediction, such as predicting an individual's next location [130, 223] or the likelihood or time duration of car accidents [19, 174, 233]. Sequential and trajectory analyses are usually used to process trajectory and traffic flow data.

4.4 Engineering Systems

Different types of engineering systems have begun to routinely apply event forecasting methods, including: 1) civil engineering, 2) electrical engineering, 3) energy engineering, and 4) other engineering domains. Despite the variety of systems in these widely different domains, the goal is essentially to predict future abnormal or failure events in order to support the system's sustainability and robustness. Both the location and time of future events are key factors for these predictions. The input features usually consist of sensing data relevant to specific engineering systems.

- **Civil engineering.** This covers various a wide range of problems in diverse urban systems, such as smart building fault adverse event prediction [21], emergency management equipment failure prediction [66], manhole event prediction [179], and other events [99].
- **Electrical engineering**. This includes teleservice systems failures [61] and unexpected events in wire electrical discharge machining operations [184].
- **Energy engineering**. Event prediction is also a hot topic in energy engineering, as such systems usually require strong robustness to handle the disturbance from the nature environments. Active research domains here include wind power ramp prediction [83], solar power ramp prediction [1], and adverse events in low carbon energy production [50].
- **Other engineering domains**. There is also active research on event prediction in other domains, such as irrigation event prediction in agricultural engineering [161] and mechanical fault prediction in mechanical engineering [197].

4.5 Cyber Systems

Here, the prediction models proposed generally focus on either network-level events or device-level events. For both types, the general goal is essentially to predict the likelihood of future system failure or attacks based on various indicators of system vulnerability. So far these two categories have essentially differed only in their inputs: the former relies on network features, including system specifications, web access logs and search queries, mismanagement symptoms, spam, phishing, and scamming activity, although some researchers are investigating the use of social media text streams to identify semantics indicating future potential attack targets of DDoS [142, 217]. For device-level events, the features of interest are usually the binary file appearance logs of machines [160, 210]. Some work has been done on micro-architectural attacks [90] by observing and proactively analyzing the observations on speculative branches, out-of-order executions and shared last level caches [188].

4.6 Political events

Political event prediction has become a very active research area in recent years, largely thanks to the popularity of social media. The most common research topics can be categorized as: 1) offline events, and 2) online activism.

4.6.1 Offline events. This includes civil unrest [171], conflicts [218], violence [28], and riots [67]. This type of research usually targets the future events' geo-location, time, and topics by leveraging the social sensors that indicate public opinions and intentions. Utilization of social media has become a popular approach for these endeavors, as social media is a source of vital information during the event development stage [171]. Specifically, many aspects are clearly visible in social media, including complaints from the public (e.g., toward the government), discussions about their intentions regarding specific political events and targets, as well as advertisements for the planned events. Due to the richness of this information, further information on future events such as the type of event [85], the anticipated participants population [171], and the event scale [84] can also be discovered in advance.

4.6.2 Online events. Due to the major impact of online media such as online forums and social media, many events such as online activism, petitions, and hoaxes in such online platform also involve strong motivations for achieving some political purpose [213]. Beyond simple detection, the prediction of various types of events have been studied in order to enable proactive intervention to sidetrack the events such as hoaxes and rumor propagation [106]. Other researchers have sought to foresee the results of future political events in order to benefit a particular group of practitioners, for example by predicting the outcome of online petitions or presidential elections [213].

4.7 Natural disasters

Different types of natural disasters have been the focus of a great deal of research. Typically, these are rare events, but mechanistic models, a long historical records (often extending back dozens or hundreds of years), and domain knowledge are usually available. The input data are typically collected by sensors or sensor networks and the output is the risk or hazard of future potential events. Since these event occurrence are typically rare but very high-stakes, many researchers strive to cover all event occurrences and hence aim to ensure high recalls.

4.7.1 Geophysics-related. **Earthquakes.** Predictions here typically focus on whether there will be an earthquake with a magnitude larger than a specified threshold in a certain area during a future period of time. To achieve this, the original sensor data is usually proccessed using geophysical models such as GutenbergâĂŞRichterâĂŹs inverse law, the distribution of characteristic earthquake

magnitudes, and seismic quiescence [14, 175]. The processed data are then input into machine learning models that treat them as input features for predicting the output, which can be either binary values of event occurrence or time-to-event values. Some studies are devoted to identifying the time of future earthquakes and their precursors, based on an ensemble of regressors and feature selection techniques [178], while others focus on aftershock prediction and the consequences of the earthquake, such as fire prevention [140]. It worth noting that social media data has also been used for such tasks, as this often supports early detection of the first-wave earthquake, which can then be used to predict the afterstocks or earthquakes in other locations [181].

Fire events. Research in this category can be grouped into urban fires and wildfires. This type of research often focuses on the time at which a fire will affect a specific location, such as a building. The goal here is to predict the risk of future fire events. To achieve this, both the external environment and the intrinsic properties of the location of interests are important. Therefore, both static input data (e.g., natural conditions and demographics) and time-varying data (e.g., weather, climate, and crowd flow) are usually involved. Shin and Kim [189] focus on building fire risk prediction, where the input is the building's profile. Others have studied wildfires, where weather data and satellite data are important inputs. This type of research focuses primarily on predicting both the time and location of future fires [216, 234].

Other researchers have focused on rarer events such as volcanic eruptions. For example, some leverage chemical prior knowledge to build a Bayesian network for prediction [44], while others adopt point processes to predict the hazard of future events [56].

4.7.2 Atmospheric science-related. **Flood events**. Floods may be caused by many different reasons, including atmospheric (e.g., snow and rain), hydrological (e.g., ice melting, wind-generated waves, and river flow), and geophysical (e.g., terrain) conditions. This makes the forecasting of floods highly complicated task that requires multiple diverse predictors [212]. Flood event prediction has a long history, with the latest research focusing especially on computational and simulation models based on domain knowledge. This usually involves using ensemble prediction systems as inputs for hydrological and/or hydraulic models to produce river discharge predictions. For a detailed survey on flood computational models please refer to [47]. However, it is prohibitively difficult to comprehensively consider and model all the factors correctly while avoiding all the accumulated errors from upstream predictions (e.g., precipitation prediction). Another direction, based on data-driven models such as statistical and machine learning models for flood prediction, is deemed promising and is expected to be complementary to existing computational models. These newly developed machine learning models are often based solely on historical data, requiring no knowledge of the underlying physical processes. Representative models are SVM, random forests, and neural networks and their variants and hybrids. A detailed recent survey is provided in [149].

Tornado Forecasting. Tornadoes usually develop within thunderstorms and hence most tornado warning systems are based on the prediction of thunderstorms. For a comprehensive survey, please refer to [68]. Machine learning models, when applied to tornado forecasting tasks, usually suffer from high-dimensionality issues, which are very common in meteorological data. Some methods have leveraged dimensional reduction strategies to preprocess the data [230] before prediction. Research on other atmosphere-related events such as droughts and ozone events has also been conducted [77].

4.7.3 Astrophysics-related. There is a large body of prediction research focusing on events outside the Earth, especially those affecting the star closest to us, the sun. Methods have been proposed to predict various solar events that could impact life on Earth, including solar flares [20], solar eruptions [4], and high energy particle storms [141]. The goal here is typically to use satellite

imagery data of the sun to predict the time and location of future solar events and the activity strength [74].

4.8 Business

Business intelligence can be grouped into company-based events and customer-based events.

4.8.1 Customer activity prediction. The most important customer activities in business is whether a customer will continue doing business with a company and how long a costumer will be willing to wait before receiving the service? A great deal of research has been devoted to these topics, which can be categorized based on the type of business entities namely enterprises, social media, and education, who are primarily interested in churn prediction, site migration, and student dropout, respectively. The first of these focuses on predicting whether and when a customer is likely to stop doing business with a profitable enterprise [71]. The second aims to predict whether a social media user will move from one site, such as Flickr, to another, such as Instagram, a movement known as site migration [236]. While site migration is not popular, attention migration might actually be much more common, as a user may "move" their major activities from one social media site to another. The third type, student dropout, is a critical domain for education data mining, where the goal is to predict the occurrence of absenteeism from school for no good reason for a continuous number of days; a comprehensive survey is available in [143]. For all three types, the procedure is first to collect features of a customer's profile and activities over a period of time and then conventional or sequential classifiers or regressors are generally used to predict the occurrence or time-to-event of the future targeted activity.

4.8.2 Business process events. Financial event prediction has been attracting a huge amount of attention for risk management, marketing, investment prediction and fraud prevention. Multiple information resources, including news, company announcements, and social media data could be utilized as the input, often taking the form of time series or temporal sequences. These sequential inputs are used for the prediction of the time and occurrence of future high-stack events such as company distress, suspension, mergers, dividends, layoffs, bankruptcy, and market trends (rises and falls in the company's stock price) [36, 40, 65, 91, 122, 144, 226].

4.9 Crime

It is difficult to deduce the precise location and time for individual crime incidences. Therefore, the focus is instead estimating the risk and probability of the location, time, and types of future crimes. This field can be naturally categorized based on the various crime types:

4.9.1 Political crimes and terrorism. This type of crime is typically highly destructive, and hence attracts huge attention in its anticipation and prevention. Terrorist activities are usually aimed at religious, political, iconic, economic or social targets. The attacker typically targets larger numbers of people and the evidences related to such attacks is retained in the long run. Though it is extremely challenging to predict the precise location and time of individual terrorism incidents, numerous studies have shown the potential utility for predicting the regional risks of terrorism attacks based on information gathered from many data sources such as geopolitical data, weather data, and economics data. The Global Terrorism Database is the most widely recognized dataset that records the descriptions of world-wide terrorism events of recent decades. In addition to terrorism events, other similar events such as mass killings [202] and armed-conflict events [193] have also been studied using similar problem formulations.

4.9.2 Crime incidents. Most studies on this topic focus on predicting the types, intensity, count, and probability of crime events across defined geo-spatial regions. Until now, urban crimes are

most commonly the topic of research due to data availability. The geospatial characteristics of the urban areas, their demographics, and temporal data such as news, weather, economics, and social media data are usually used as inputs. The geospatial dependency and correlation of the crime patterns are usually leveraged during the prediction process using techniques originally developed for spatial predictions, such as kernel density estimation and conditional random fields. Some works simplify the tasks by only focusing on specific types of crimes such as theft [180], robbery, and burglary [51].

4.9.3 Organized and serial crimes. Unlike the above research on regional crime risks, some recent studies strive to predict the next incidents of criminal individuals or groups. This is because different offenders may demonstrate different behavioral patterns, such as targeting specific regions (e.g., wealthy neighborhoods), victims (e.g., women), for specific benefits (e,g, money). The goal here is thus is to predict the next crime site and/or time, based on the historical crime event sequence of the targeted criminal individual or group. Models such as point processes [130] or Bayesian networks [133] are usually used to address such problems.

5 OPEN CHALLENGES AND FUTURE DIRECTIONS

Despite the major advances in event prediction in recent years, there are still a number of open problems and potentially fruitful directions for future research, as follows:

5.1 Model Transparency, Interpretability, and Accountability

Increasingly sophisticated forecasting models have been proposed to improve the prediction accuracy, including those utilizing approaches such as ensemble models, neural networks, and the other complex systems mentioned above. However, although the accuracy can be improved, the event prediction models are rapidly becoming too complex to be interpreted by human operators. The need for better model accountability and interpretability is becoming an important issue; as big data and Artificial Intelligence techniques are applied to ever more domains this can lead to serious consequences for applications such as healthcare and disaster management. Models that are not interpretable by humans will find it hard to build the trust needed if they are to be fully integrated into the workflow of practitioners. A closely related key feature is the accountability of the event prediction system. For example, disaster managers need to thoroughly understand a model's recommendations if they are to be able to explain the reason for a decision to displace people in a court of law. Moreover, an ever increasing number of laws in countries around the world are beginning to require adequate explanations of decisions reached based on model recommendations. For example, Articles 13-15 in the European Union's General Data Protection Regulation (GDPR) [207] require algorithms that make decisions that âĂIJsignificantly affect" individuals to provide explanations ("right to explanation") by May 28, 2018. Similar laws have also been established in countries such as the United States [48] and China [166].

5.2 Vulnerability to Noise and Adversarial Attacks

The massive popularity of the proposal, development, and deployment of event prediction is stimulating a surge interest in developing ways to counter-attack these systems. It will therefore not be a surprise when we begin to see the introduction of techniques to obfuscate these event prediction methods in the near future. As with many state-of-the-art AI techniques applied in other domains such as object recognition, event prediction methods can also be very vulnerable to noise and adversarial attacks. The famous failure of Google Flu trends, which missed the peak of the 2013 flu season by 140 percent due to low relevance and high disturbance affecting the input signal, is a vivid memory for practitioners in the field [82]. Many predictions relying on social media data

can also be easily influenced or flipped by injecting scam messages. Event prediction models also tend to over-rely on low-quality input data that can be easily disturbed or manipulated, lacking sufficient robustness to survive noisy signals and adversarial attacks. Similar problems threaten to other application domains such as business intelligence, crime, and cyber systems.

5.3 Integration of mechanistic knowledge and data driven-models

Over the years, many domains have accumulated a significant amount of knowledge and experience about event development occurrence mechanisms, which can thus provide important clues for anticipating future events, such as epidiomiology models, socio-political models, and earthquake models. All of these models focus on simplifying real-world phenomena into concise principles in order to grasp the core mechanism, discarding many details in the process. In contrast, data-driven models strive to ensure the accurate fitting of large historical data sets, based on sufficient model expressiveness but cannot guarantee that the true underlying principle and causality of event occurrence modeled accurately. There is thus a clear motivation to combine their complementary strengths, and although this has already attracted great deal of interest [98, 242], most of the models proposed so far are merely ensemble learning-based and simply merge the final predictions from each model. A more thorough integration is needed that can directly embed the core principles to regularize and instruct the training of data-driven event prediction methods. Moreover, existing attempts are typically specific to particular domains and are thus difficult to develop further as they require in-depth collaborations between data scientists and domain experts. A generic framework developed to encompass multiple different domains is imperative and would be highly beneficial for the various domain experts.

5.4 Prescriptive and counterfactual analysis

The ultimate purpose of event prediction is usually not just to anticipate the future, but to change it, for example by avoiding a system failure and flattening the curve of a disease outbreak. However, it is difficult for practitioners to determine how to act appropriately and implement effective policies in order to achieve the desired results in the future. This requires a capability that goes beyond simply predicting future events based on the current situation, requiring them instead to also take into account the new actions being taken in real time and then predict how they might influence the future. One promising direction is the use of counterfactual event [145] prediction that models what would have happened if different circumstances had occurred. Another related direction is prescriptive analysis where different actions can be merged into the prediction system and future results anticipated or optimized. Related works have been developed in few domains such as epidemiology. However, as yet these lack sufficient research in many other domains that will be needed if we are to develop generic frameworks that can benefit different domains.

5.5 Multi-objective training

Existing event prediction methods mostly focus primarily on accuracy. However, decision makers who utilize these predicted event results usually need much more, including key factors such as event resolution (e.g., time resolution, location resolution, description details), confidence (e.g., the probability a predicted event will occur), efficiency (whether the model can predict per day or per seccond), lead time (how many days the prediction can be made prior to the event occurring), and event intensity (how serious it is). multi-objective optimization (e.g., accuracy, confidence, resolution). There are typically trade-offs among all the above metrics and accuracy, so merely optimizing accuracy during training will inevitably mean the results drift away from the overall optimal event-prediction-based decision. A system that can flexibly balance the trade-off between

these metrics based on decision makers' needs and achieve a multi-objective optimization is the ultimate objective for these models.

6 CONCLUSION

This survey has presented a comprehensive survey of existing methodologies developed for event prediction methods in the big data era. It provides an extensive overview of the event prediction challenges, techniques, applications, evaluation procedures, and future outlook, summarizing the research presented in over 200 publications, most of which were published in the last five years. Event prediction challenges, opportunities, and formulations have been discussed in terms of the event element to be predicted, including the event location, time, and semantics, after which we went on to propose a systematic taxonomy of the existing event prediction techniques according to the formulated problems and types of methodologies designed for the corresponding problems. We have also analyzed the relationships, differences, advantages, and disadvantages of these techniques from various domains, including machine learning, data mining, pattern recognition, natural language processing, information retrieval, statistics, and other computational models. In addition, a comprehensive and hierarchical categorization of popular event prediction applications has been provided that covers domains ranging from natural science to the social sciences. Based upon the numerous historical and state-of-the-art works discussed in this survey, the paper concludes by discussing open problems and future trends in this fast-growing domain.

REFERENCES

[1] Mohamed Abuella and Badrul Chowdhury. 2019. Forecasting of solar power ramp events: A post-processing approach. *Renewable Energy* 133 (2019), 1380–1392.

[2] Saurav Acharya, Byung Suk Lee, and Paul Hines. 2017. Causal prediction of top-k event types over real-time event streams. *Comput. J.* 60, 11 (2017), 1561–1581.

[3] Bijaya Adhikari, Xinfeng Xu, Naren Ramakrishnan, and B Aditya Prakash. 2019. Epideep: Exploiting embeddings for epidemic forecasting. In *SIGKDD 2019*. 577–586.

[4] Ashna Aggarwal, Nicole Schanche, Katharine K Reeves, Dustin Kempton, and Rafal Angryk. 2018. Prediction of solar eruptions using filament metadata. *The Astrophysical Journal Supplement Series* 236, 1 (2018), 15.

[5] Mohammad Al Boni and Matthew S Gerber. 2016. Area-specific crime prediction models. In *ICMLA 2016*. IEEE, 671–676.

[6] Hafiz A Alaka, Lukumon O Oyedele, Hakeem A Owolabi, et al. 2016. Methodological approach of construction business failure prediction studies: a review. *Construction Management and Economics* 34, 11 (2016), 808–842.

[7] Elias Alevizos, Alexander Artikis, and George Paliouras. 2017. Event forecasting with pattern Markov chains. In *Proceedings of the 11th ACM International Conference on Distributed and Event-based Systems*. 146–157.

[8] Elias Alevizos, Alexander Artikis, and Georgios Paliouras. 2018. Wayeb: A tool for complex event forecasting. *arXiv preprint arXiv:1901.01826* (2018).

[9] Elias Alevizos, Anastasios Skarlatidis, Alexander Artikis, and Georgios Paliouras. 2017. Probabilistic complex event recognition: A survey. *ACM Computing Surveys (CSUR)* 50, 5 (2017), 1–31.

[10] James Allan, Ron Papka, and Victor Lavrenko. [n.d.]. On-line new event detection and tracking. In *SIGIR 1998*. 37–45.

[11] M Antunes, MA Amaral Turkman, and KF Turkman. 2003. A Bayesian approach to event prediction. *Journal of Time Series Analysis* 24, 6 (2003), 631–646.

[12] Marta Arias, Argimiro Arratia, and Ramon Xuriguera. 2014. Forecasting with Twitter data. *TIST* 5, 1 (2014), 1–24.

[13] Jason Asher. 2018. Forecasting Ebola with a regression transmission model. *Epidemics* 22 (2018), 50–55.

[14] KM Asim, F Martínez-Álvarez, A Basit, and T Iqbal. 2017. Earthquake magnitude prediction in Hindukush region using machine learning techniques. *Natural Hazards* 85, 1 (2017), 471–486.

[15] Farzindar Atefeh and Wael Khreich. 2015. A survey of techniques for event detection in Twitter. *Computational Intelligence* 31, 1 (2015), 132–164.

[16] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al. 1999. *Modern information retrieval*. Vol. 463. ACM press New York.

[17] Gökhan BakIr, Thomas Hofmann, Bernhard Schölkopf, Alexander J Smola, and Ben Taskar. 2007. *Predicting structured data*. MIT press.

[18] Michel Ballings and Dirk Van den Poel. 2012. Customer event history for churn prediction: How long is long enough? *Expert Systems with Applications* 39, 18 (2012), 13517–13522.

[19] Jie Bao, Pan Liu, and Satish V Ukkusuri. 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. *Accident Analysis & Prevention* 122 (2019), 239–254.
[20] Graham Barnes, KD Leka, CJ Schrijver, et al. 2016. A comparison of flare forecasting methods. I. Results from the âĂIJall-clearâĂİ workshop. *The Astrophysical Journal* 829, 2 (2016), 89.
[21] Aniruddha Basak, Ole Mengshoel, Stefan Hosein, and Rodney Martin. 2016. Scalable causal learning for predicting adverse events in smart buildings. In *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*.
[22] Hila Becker, Dan Iter, Mor Naaman, and Luis Gravano. 2012. Identifying content for planned events across social media sites. In *WSDM 2012*. 533–542.
[23] Juan-Jose Beunza, Enrique Puertas, Ester García-Ovejero, et al. 2019. Comparison of machine learning algorithms for clinical event prediction. *Journal of Biomedical Informatics* 97 (2019), 103257.
[24] Stephan Bialonski, Gerrit Ansmann, and Holger Kantz. 2015. Data-driven prediction and prevention of extreme events in a spatially extended excitable system. *Physical Review E* 92, 4 (2015), 042910.
[25] Marin Biloš, Bertrand Charpentier, and Stephan Günnemann. 2019. Uncertainty on Asynchronous Time Event Prediction. In *NeurIPS 2019*. 12831–12840.
[26] Christopher M Bishop. 2006. *Pattern recognition and machine learning*. Springer.
[27] Keith R Bisset, Jiangzhuo Chen, Xizhou Feng, VS Anil Kumar, and Madhav V Marathe. 2009. EpiFast: A fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In *SC 2009*. 430–439.
[28] Robert A Blair, Christopher Blattman, and Alexandra Hartman. 2017. Predicting local violence: Evidence from a panel survey in Liberia. *Journal of Peace Research* 54, 2 (2017), 298–312.
[29] Robert A Blair and Nicholas Sambanis. 2020. Forecasting civil wars: theory and structure in an age of "big data" and machine learning. *Journal of Conflict Resolution* (2020), 0022002720918923.
[30] Patrick T Brandt, John R Freeman, and Philip A Schrodt. 2011. Real time, time series forecasting of inter-and intra-state political conflict. *Conflict Management and Peace Science* 28, 1 (2011), 41–64.
[31] Jose Cadena, Gizem Korkmaz, Chris J Kuhlman, Achla Marathe, Naren Ramakrishnan, and Anil Vullikanti. 2015. Forecasting social unrest using activity cascades. *PloS one* 10, 6 (2015), e0128879.
[32] Raffaella Calabrese and Johan A Elkink. 2016. Estimating binary spatial autoregressive models for rare events. In *Advances in Econometrics, Vol. 37*. 145–166.
[33] Flávia Dias Casagrande, Jim Tørresen, and Evi Zouganeli. 2018. Sensor event prediction using recurrent neural network in smart homes for older adults. In *International Conference on Intelligent Systems 2018*. IEEE, 662–668.
[34] Flávia Dias Casagrande and Evi Zouganeli. 2019. Prediction of next sensor event and its time of occurrence using transfer learning across homes. In *International Conference on Computer Systems and Applications*. IEEE, 1–8.
[35] Finneas JR Catling and Anthony H Wolff. 2020. Temporal convolutional networks allow early prediction of events in critical care. *Journal of the American Medical Informatics Association* 27, 3 (2020), 355–365.
[36] Mark Cecchini, Haldun Aytug, Gary J Koehler, and Praveen Pathak. 2010. Making words work: Using financial text as a predictor of financial events. *Decision Support Systems* 50, 1 (2010), 164–175.
[37] Pew Reseach Center. 2017. Social media fact sheet. *Pew Research Center: Washington, DC, USA* (2017).
[38] Deepayan Chakrabarti and Kunal Punera. 2011. Event summarization using tweets. In *ICWSM 2011*.
[39] Ki Chan and Wai Lam. 2005. Extracting causation knowledge from natural language texts. *International Journal of Intelligent Systems* 20, 3 (2005), 327–358.
[40] Samuel WK Chan and James Franklin. 2011. A text-based decision support system for financial sequence prediction. *Decision Support Systems* 52, 1 (2011), 189–198.
[41] Feng Chen and Daniel B Neill. 2014. Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In *SIGKDD 2014*. 1166–1175.
[42] Feng Chen, Baojian Zhou, Adil Alim, and Liang Zhao. 2017. A generic framework for interesting subspace cluster detection in multi-attributed networks. In *ICDM 2017*. IEEE, 41–50.
[43] Meng Chen, Xiaohui Yu, and Yang Liu. 2018. PCNN: Deep convolutional networks for short-term traffic congestion prediction. *IEEE Transactions on Intelligent Transportation Systems* 19, 11 (2018), 3550–3559.
[44] Seong-Pyo Cheon, Sungshin Kim, So-Young Lee, and Chong-Bum Lee. 2009. Bayesian networks based rare event prediction with sensor data. *Knowledge-Based Systems* 22, 5 (2009), 336–343.
[45] Chung-Wen Cho, Ying Zheng, Yi-Hung Wu, and Arbee LP Chen. 2008. A tree-based approach for event prediction using episode rules over event streams. In *International Conference on Database and Expert Systems Applications*. Springer, 225–240.
[46] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. 2018. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *CVPR 2018*. 8789–8797.
[47] HL Cloke and Florian Pappenberger. 2009. Ensemble flood forecasting: A review. *Journal of Hydrology* 375, 3-4 (2009), 613–626.

[48] Cary Coglianese and David Lehr. 2016. Regulating by robot: Administrative decision making in the machine-learning era. *Geo. LJ* 105 (2016), 1147.
[49] Ryan Compton, Craig Lee, Jiejun Xu, Luis Artieda-Moncada, Tsai-Ching Lu, Lalindra De Silva, and Michael Macy. 2014. Using publicly visible social media to build detailed forecasts of civil unrest. *Security informatics* 3, 1 (2014), 4.
[50] Stefano Coniglio, Anthony J Dunn, and Alain B Zemkoho. 2020. Infrequent adverse event prediction in low carbon energy production using machine learning. *arXiv preprint arXiv:2001.06916* (2020).
[51] Bitzel Cortez, Berny Carrera, Young-Jin Kim, and Jae-Yoon Jung. 2018. An architecture for emergency event prediction using LSTM recurrent neural networks. *Expert Systems with Applications* 97 (2018), 315–324.
[52] Meggan E Craft, Erik Volz, Craig Packer, and Lauren Ancel Meyers. 2011. Disease transmission in territorial populations: The small-world network of Serengeti lions. *Journal of the Royal Society Interface* 8, 59 (2011), 776–786.
[53] Attila Csenki. 1990. Bayes predictive analysis of a fundamental software reliability model. *IEEE Transactions on Reliability* 39, 2 (1990), 177–183.
[54] Alessandro Daidone, Felicita Di Giandomenico, Andrea Bondavalli, and Silvano Chiaradonna. 2006. Hidden Markov models as a support for diagnosis: Formalization of the problem and synthesis of the solution. In *25th IEEE Symposium on Reliable Distributed Systems*. IEEE, 245–256.
[55] Robert Dale. 2018. Text analytics APIs, part 2: The smaller players. *Natural Language Engineering* 24, 5 (2018), 797–803.
[56] Magret Damaschke, Shane J Cronin, and Mark S Bebbington. 2018. A volcanic event forecasting model for multiple tephra records, demonstrated on Mt. Taranaki, New Zealand. *Bulletin of Volcanology* 80, 1 (2018), 9.
[57] Sina Dami, Ahmad Abdollahzadeh Barforoush, and Hossein Shirazi. 2018. News events prediction using Markov logic networks. *Journal of Information Science* 44, 1 (2018), 91–109.
[58] Arno De Caigny, Kristof Coussement, and Koen De Bock. 2018. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research* 269, 2 (2018), 760–772.
[59] Arno De Caigny, Kristof Coussement, and Koen De Bock. 2020. Leveraging fine-grained transaction data for customer life event predictions. *Decision Support Systems* 130 (2020), 113232.
[60] Tom Decroos, Vladimir Dzyuba, Jan Van Haaren, and Jesse Davis. 2017. Predicting soccer highlights from spatio-temporal match event streams. In *AAAI 2017*.
[61] Akash Deep, Dharmaraj Veeramani, and Shiyu Zhou. 2019. Event Prediction for Individual Unit Based on Recurrent Event Data Collected in Teleservice Systems. *IEEE Transactions on Reliability* 69, 1 (2019), 216–227.
[62] Walter H. Dempsey, Alexander Moreno, Christy K. Scott, et al. 2017. iSurvive: An interpretable, event-time prediction model for mHealth. In *ICML 2017*. 970–979.
[63] Jingsheng Deng, Fengcai Qiao, et al. 2015. An overview of event extraction from Twitter. In *2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. IEEE, 251–256.
[64] Xiaolei Di, Yu Xiao, Chao Zhu, Yang Deng, Qinpei Zhao, and Weixiong Rao. 2019. Traffic congestion prediction by spatiotemporal propagation patterns. In *MDM 2019*. IEEE, 298–303.
[65] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *IJCAI 2015*. 2327–2333.
[66] Zuohua Ding, Yuan Zhou, Geguang Pu, and MengChu Zhou. 2018. Online failure prediction for railway transportation systems based on fuzzy rules and data analysis. *IEEE Transactions on Reliability* 67, 3 (2018), 1143–1158.
[67] Raimundo Dos Santos, Sumit Shah, Feng Chen, Arnold Boedihardjo, Chang-Tien Lu, and Naren Ramakrishnan. 2014. Forecasting location-based events with spatio-temporal storytelling. In *SIGSPATIAL 2014*. 13–22.
[68] Charles A Doswell III, Steven J Weiss, and Robert H Johns. 1993. Tornado forecasting: A review. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophys. Monogr* 79 (1993), 557–571.
[69] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. 2016. Recurrent marked temporal point processes: Embedding event history to vector. In *SIGKDD 2016*. 1555–1564.
[70] Huilong Duan, Zhoujian Sun, Wei Dong, et al. 2019. On clinical event prediction in patient treatment trajectory using longitudinal electronic health records. *IEEE Journal of Biomedical and Health Informatics* (2019).
[71] Kamya Eria and Booma Poolan Marikannan. 2018. Systematic review of customer churn prediction in the telecom sector. *Journal of Applied Technology and Innovation* 2, 1 (2018).
[72] Şeyda Ertekin, Cynthia Rudin, Tyler H McCormick, et al. 2015. Reactive point processes: A new approach to predicting power failures in underground electrical systems. *The Annals of Applied Statistics* 9, 1 (2015), 122–144.
[73] Ali Mert Ertugrul, Yu-Ru Lin, Christina Mair, and T Taskaya Temizel. 2018. Forecasting heroin overdose occurrences from crime incidents. SBP-BRiMS 2018.
[74] David A Falconer, Ronald L Moore, Abdulnasser F Barghouty, and Igor Khazanov. 2014. MAG4 versus alternative techniques for forecasting active region flare productivity. *Space Weather* 12, 5 (2014), 306–317.
[75] Christiane Fellbaum. 2012. WordNet. *The encyclopedia of applied linguistics* (2012).

[76] Carlos Ferreira, Joao Gama, L Matias, Audun Botterud, and J Wang. 2011. *A survey on wind power ramp forecasting*. Technical Report. Argonne National Lab.(ANL), Argonne, IL (United States).
[77] Christopher B Field, Vicente Barros, Thomas F Stocker, and Qin Dahe. 2012. *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*. Cambridge University Press.
[78] Ioannis Flouris, Nikos Giatrakos, Antonios Deligiannakis, et al. 2017. Issues in complex event processing: Status and prospects in the big data era. *Journal of Systems and Software* 127 (2017), 217–236.
[79] Ilenia Fronza, Alberto Sillitti, Giancarlo Succi, Mikko Terho, and Jelena Vlasenko. 2013. Failure prediction based on log files using random indexing and support vector machines. *Journal of Systems and Software* 86, 1 (2013), 2–11.
[80] Kaiqun Fu, Taoran Ji, Liang Zhao, and Chang-Tien Lu. 2019. Titan: A spatiotemporal feature learning framework for traffic incident duration prediction. In *SIGSPATIAL 2019*. 329–338.
[81] Lajos Jenő Fülöp, Gabriella Tóth, Róbert Rácz, János Pánczél, et al. 2010. Survey on complex event processing and predictive analytics. In *Proceedings of the Fifth Balkan Conference in Informatics*. Citeseer, 26–31.
[82] Kaiser Fung. 2014. Google Flu Trends' failure shows good data> big data. *Harvard Business Review* 25 (2014), 2014.
[83] Cristobal Gallego-Castillo, Alvaro Cuerva-Tejero, and Oscar Lopez-Garcia. 2015. A review on the recent history of wind power ramp forecasting. *Renewable and Sustainable Energy Reviews* 52 (2015), 1148–1157.
[84] Yuyang Gao and Liang Zhao. 2018. Incomplete label multi-task ordinal regression for spatial event scale forecasting. In *AAAI 2018*.
[85] Yuyang Gao, Liang Zhao, Lingfei Wu, Yanfang Ye, Hui Xiong, and Chaowei Yang. 2019. Incomplete label multi-task deep learning for spatio-temporal event subtype forecasting. In *AAAI 2019*, Vol. 33. 3638–3646.
[86] M Ghil, P Yiou, Stéphane Hallegatte, BD Malamud, P Naveau, A Soloviev, V Keilis-Borok, D Kondrashov, V Kossobokov, O Mestre, et al. 2011. Extreme events: Dynamics, statistics and prediction. *Nonlinear Processes in Geophysics* 18, 3 (2011), 295.
[87] Fatma Ezzahra Gmati, Salem Chakhar, Wided Lejouad Chaari, and Mark Xu. 2019. A taxonomy of event prediction methods. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 12–26.
[88] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press.
[89] Mark Granroth-Wilding and Stephen Clark. 2016. What happens next? event prediction using a compositional neural network model. In *AAAI 2016*. 2727âĂŞ2733.
[90] Berk Gulmezoglu, Ahmad Moghimi, Thomas Eisenbarth, and Berk Sunar. 2019. Fortuneteller: Predicting microarchitectural attacks via unsupervised deep learning. *arXiv preprint arXiv:1907.03651* (2019).
[91] Michael Hagenau, Michael Liebmann, Markus Hedwig, and Dirk Neumann. 2012. Automated news reading: Stock price prediction based on financial news using context-specific features. In *2012 45th Hawaii International Conference on System Sciences*. IEEE, 1040–1049.
[92] Jiawei Han, Jian Pei, and Micheline Kamber. 2011. *Data mining: concepts and techniques*. Elsevier.
[93] Mengmeng Hao, Dong Jiang, Fangyu Ding, Jingying Fu, and Shuai Chen. 2019. Simulating spatio-temporal patterns of terrorism incidents on the Indochina Peninsula with GIS and the random forest method. *IJGI* 8, 3 (2019), 133.
[94] Chikara Hashimoto, Kentaro Torisawa, Julien Kloetzer, et al. 2014. Toward future scenario generation: Extracting event causality exploiting semantic relation, context, and association features. In *ACL 2014*. 987–997.
[95] Andrew Hoegh, Scotland Leman, Parang Saraf, and Naren Ramakrishnan. 2015. Bayesian model fusion for forecasting civil unrest. *Technometrics* 57, 3 (2015), 332–340.
[96] Linmei Hu. 2019. Integrating hierarchical attentions for future subevent prediction. *IEEE Access* 8 (2019), 3106–3114.
[97] Linmei Hu, Juanzi Li, Liqiang Nie, Xiao-Li Li, and Chao Shao. 2017. What happens next? future subevent prediction using contextual hierarchical LSTM. In *AAAI 2017*.
[98] Ting Hua, Chandan K Reddy, Lei Zhang, Lijing Wang, Liang Zhao, Chang-Tien Lu, and Naren Ramakrishnan. 2018. Social media based simulation models for understanding disease dynamics.. In *IJCAI 2018*. 3797–3804.
[99] Chao Huang, Chuxu Zhang, Jiashu Zhao, Xian Wu, Dawei Yin, and Nitesh Chawla. 2019. Mist: A multiview and multimodal spatial-temporal learning framework for citywide abnormal event forecasting. In *WWW 2019*. 717–728.
[100] Gordon F Hughes, Joseph F Murray, Kenneth Kreutz-Delgado, and Charles Elkan. 2002. Improved disk-drive failure warnings. *IEEE Transactions on Reliability* 51, 3 (2002), 350–357.
[101] Ali HürriyetoÇğlu, Nelleke Oostdijk, and Antal van den Bosch. 2018. Estimating time to event of future events based on linguistic cues on Twitter. In *Intelligent natural language processing: Trends and applications*. 67–97.
[102] Fadil Inceoglu, Jacob H Jeppesen, Peter Kongstad, et al. 2018. Using machine learning methods to forecast if solar flares will be associated with CMEs and SEPs. *The Astrophysical Journal* 861, 2 (2018), 128.
[103] Bram Jans, Steven Bethard, Ivan Vulić, and Marie Francine Moens. 2012. Skip n-grams and ranking functions for predicting script events. In *EACL 2012*. 336–344.

[104] Renhe Jiang, Xuan Song, Dou Huang, Xiaoya Song, Tianqi Xia, et al. 2019. Deepurbanevent: A system for predicting citywide crowd dynamics at big events. In *SIGKDD 2019*. 2114–2122.
[105] Zhe Jiang. 2018. A survey on spatial prediction methods. *TKDE* 31, 9 (2018), 1645–1664.
[106] Fang Jin, Edward Dougherty, Parang Saraf, Yang Cao, and Naren Ramakrishnan. 2013. Epidemiological modeling of news and rumors on Twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*. 1–9.
[107] David Jurgens. 2013. That's what friends are for: Inferring location in online social media platforms based on social relationships. In *ICWSM 2013*.
[108] Wei Kang, Jie Chen, Jiuyong Li, Jixue Liu, Lin Liu, Grant Osborne, Nick Lothian, Brenton Cooper, Terry Moschou, and Grant Neale. 2017. Carbon: Forecasting civil unrest events by monitoring news and social media. In *International Conference on Advanced Data Mining and Applications*. Springer, 859–865.
[109] Ahmed Kattan, Shaheen Fatima, and Muhammad Arif. 2015. Time-series event-based prediction: An unsupervised learning framework based on genetic programming. *Information Sciences* 301 (2015), 99–123.
[110] Amin Vahedian Khezerlou, Xun Zhou, Ling Tong, Yanhua Li, and Jun Luo. 2019. Forecasting gathering events through trajectory destination prediction: A dynamic hybrid model. *TKDE* (2019).
[111] Christopher SG Khoo, Syin Chan, and Yun Niu. 2000. Extracting causal knowledge from a medical database using graphical patterns. In *ACL 2000*. 336–343.
[112] Jaegwon Kim. 1993. *Supervenience and mind: Selected philosophical essays*. Cambridge University Press.
[113] Hirokazu Kiyomaru, Kazumasa Omura, Yugo Murawaki, Daisuke Kawahara, and Sadao Kurohashi. 2019. Diversity-aware event prediction based on a conditional variational autoencoder with reconstruction. In *Proceedings of the First Workshop on Commonsense Inference in Natural Language Processing*. 113–122.
[114] Jack PC Kleijnen and Wim CM van Beers. 2020. Prediction for big data through Kriging: small sequential and one-shot designs. *American Journal of Mathematical and Management Sciences* (2020), 1–15.
[115] Canasai Kruengkrai, Kentaro Torisawa, et al. 2017. Improving event causality recognition with multiple background knowledge sources using multi-column convolutional neural networks. In *AAAI 2017*. 3466âĂŞ3473.
[116] Martin Kulldorff. 1997. A spatial scan statistic. *Communications in Statistics-Theory and Methods* 26, 6 (1997), 1481–1496.
[117] Florian A Kunneman and APJ van den Bosch. 2012. Leveraging unscheduled event prediction through mining scheduled event tweets. (2012).
[118] Matthew Kupilik and Frank Witmer. 2018. Spatio-temporal violent event prediction using Gaussian process regression. *Journal of Computational Social Science* 1, 2 (2018), 437–451.
[119] Håvard Kvamme, Ørnulf Borgan, and Ida Scheel. 2019. Time-to-event prediction with neural networks and Cox regression. *JMLR* 20, 129 (2019), 1–30.
[120] Daniel T Larose. 2015. *Data mining and predictive analytics*. John Wiley & Sons.
[121] Srivatsan Laxman, Vikram Tankasali, and Ryen W White. 2008. Stream prediction using a generative model based on frequent episodes in event sequences. In *SIGKDD 2008*. 453–461.
[122] Mai Le, Bogdan Gabrys, and Detlef Nauck. 2012. A hybrid model for business process event prediction. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Springer, 179–192.
[123] Lei Lei, Xuguang Ren, Nigel Franciscus, Junhu Wang, and Bela Stantic. 2019. Event Prediction Based on Causality Reasoning. In *Asian Conference on Intelligent Information and Database Systems*. Springer, 165–176.
[124] Benjamin Letham, Cynthia Rudin, et al. 2015. Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics* 9, 3 (2015), 1350–1371.
[125] Benjamin Letham, Cynthia Rudin, and David Madigan. 2013. Sequential event prediction. *Machine learning* 93, 2-3 (2013), 357–380.
[126] Eldon Y Li, Chen-Yuan Tung, and Shu-Hsun Chang. 2016. The wisdom of crowds in action: Forecasting epidemic diseases with a web-based prediction market system. *International Journal of Medical Informatics* 92 (2016), 35–43.
[127] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P Trevino, Jiliang Tang, and Huan Liu. 2017. Feature selection: A data perspective. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 1–45.
[128] Shengzhi Li, Jianzhong Qiao, and Shukuan Lin. 2017. Multi-attribute event modeling and prediction over event streams from sensors. In *International Conference on Parallel and Distributed Systems*. IEEE, 796–799.
[129] Yang Li, Nan Du, and Samy Bengio. 2017. Time-dependent representation for neural event sequence prediction. *arXiv preprint arXiv:1708.00065* (2017).
[130] Yunyi Li and Tong Wang. [n.d.]. Next hit predictor-self-exciting risk modeling for predicting next locations of serial crimes. *NIPS Workshop on AI for Social Good, 2018* ([n. d.]).
[131] Zhongyang Li, Xiao Ding, and Ting Liu. 2018. Constructing narrative event evolutionary graph for script event prediction. *arXiv preprint arXiv:1805.05081* (2018).
[132] Zhiguo Li, Shiyu Zhou, Suresh Choubey, and Crispian Sievenpiper. 2007. Failure event prediction using the Cox proportional hazard model driven by frequent failure signatures. *IIE transactions* 39, 3 (2007), 303–315.

[133] Renjie Liao, Xueyao Wang, Lun Li, and Zengchang Qin. 2010. A novel serial crime prediction model based on Bayesian learning theory. In *International Conference on Machine Learning and Cybernetics*. 1757–1762.
[134] Li Lin, Lijie Wen, and Jianmin Wang. 2019. Mm-pred: A deep predictive model for multi-attribute event sequence. In *ICDM 2019*. SIAM, 118–126.
[135] Ying-Lung Lin, Meng-Feng Yen, and Liang-Chih Yu. 2018. Grid-based crime prediction using geographical features. *ISPRS International Journal of Geo-Information* 7, 8 (2018), 298.
[136] Hua Liu and Donald E Brown. 2004. A new point process transition density model for space-time event prediction. *IEEE Transactions on Systems, Man, and Cybernetics, Part C* 34, 3 (2004), 310–324.
[137] Hugo Liu and Push Singh. 2004. ConceptNetâĂŤa practical commonsense reasoning tool-kit. *BT Technology Journal* 22, 4 (2004), 211–226.
[138] Jiasen Lu, Caiming Xiong, Devi Parikh, and Richard Socher. 2017. Knowing when to look: Adaptive attention via a visual sentinel for image captioning. In *CVPR 2017*. 375–383.
[139] Shangwen Lv, Wanhui Qian, Longtao Huang, Jizhong Han, and Songlin Hu. 2019. SAM-Net: Integrating event-level and chain-level attentions to predict what happens next. In *AAAI 2019*, Vol. 33. 6802–6809.
[140] Roxane Mallouhy, Chady Abou Jaoude, Christophe Guyeux, and Abdallah Makhoul. 2019. Major earthquake event prediction using various machine learning algorithms. In *ICT-DM 2019*. IEEE, 1–7.
[141] Petrus C Martens and Rafal A Angryk. 2017. Data handling and assimilation for solar event prediction. *Proceedings of the International Astronomical Union* 13, S335 (2017), 344–347.
[142] Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos, Tomoharu Iwata, and Masatoshi Yoshikawa. 2012. Fast mining and forecasting of complex time-stamped events. In *SIGKDD 2012*. 271–279.
[143] Neema Mduma, Khamisi Kalegele, and Dina Machuve. 2019. A survey of machine learning approaches and techniques for student dropout prediction. *Data Science Journal* 18, 1 (2019).
[144] Nijat Mehdiyev, Joerg Evermann, and Peter Fettke. 2017. A multi-stage deep learning approach for business process event prediction. In *2017 IEEE 19th Conference on Business Informatics (CBI)*, Vol. 1. IEEE, 119–128.
[145] Peter Menzies. 2001. Counterfactual theories of causation. In *The Stanford Encyclopedia of Philosophy*. Stanford University, 1–7.
[146] Evangelos Michelioudakis, Anastasios Skarlatidis, Elias Alevizos, et al. 2016. Event recognition and forecasting technology. *Technical Report* (2016).
[147] Bryan Minor and Diane Cook. 2017. Forecasting occurrences of activities. *Pervasive and Mobile Computing* 38 (2017), 77 – 91. https://doi.org/10.1016/j.pmcj.2016.09.010
[148] Mehrnoosh Mirtaheri, Sami Abu-El-Haija, Tozammel Hossain, et al. [n.d.]. Tensor-based method for temporal geopolitical event forecasting. *ICML Workshop on Learning and Reasoning with Graph-Structured Data, 2019* ([n. d.]).
[149] Amir Mosavi, Pinar Ozturk, and Kwok-wing Chau. 2018. Flood prediction using machine learning models: Literature review. *Water* 10, 11 (2018), 1536.
[150] Ksenia D Mukhina, Alexander A Visheratin, and Denis Nasonov. 2019. Urban events prediction via convolutional neural networks and Instagram data. *Procedia Computer Science* 156 (2019), 176–184.
[151] Sathappan Muthiah, Patrick Butler, Rupinder Paul Khandpur, Parang Saraf, et al. 2016. EMBERS at 4 years: Experiences operating an open source indicators forecasting system. In *SIGKDD 2016*. 205–214.
[152] Sathappan Muthiah, Bert Huang, Jaime Arredondo, David Mares, Lise Getoor, Graham Katz, and Naren Ramakrishnan. 2016. Capturing planned protests from open source indicators. *AI Magazine* 37, 2 (2016), 63–75.
[153] Yoko Nakajima, Michal Ptaszynski, F Masui, and H Hirotoshi. 2017. A Prototype Method for Future Event Prediction Based on Future Reference Sentence Extraction. In *IJCAI 2017*. 42–49.
[154] Lukas Neumann, Andrew Zisserman, and Andrea Vedaldi. 2019. Future event prediction: If and when. In *CVPR Workshops 2019*.
[155] Dai Quoc Nguyen, Dat Quoc Nguyen, Cuong Xuan Chu, Stefan Thater, and Manfred Pinkal. 2017. Sequence to sequence learning for event prediction. *arXiv preprint arXiv:1709.06033* (2017).
[156] Yue Ning, Rongrong Tao, Chandan K Reddy, Huzefa Rangwala, James C Starz, and Naren Ramakrishnan. 2018. STAPLE: Spatio-temporal precursor learning for event forecasting. In *SDM 2018*. SIAM, 99–107.
[157] Yue Ning, Liang Zhao, Feng Chen, Chang-Tien Lu, and Huzefa Rangwala. 2019. Spatio-temporal event forecasting and precursor identification. In *SIGKDD 2019*. 3237–3238.
[158] Hiroshi Nishiura. 2011. Real-time forecasting of an epidemic using a discrete time stochastic model: a case study of pandemic influenza (H1N1-2009). *Biomedical Engineering Online* 10, 1 (2011), 15.
[159] Maya Okawa, Tomoharu Iwata, Takeshi Kurashima, Yusuke Tanaka, et al. 2019. Deep mixture point processes: Spatio-temporal event prediction with rich contextual information. In *SIGKDD 2019*. 373–383.
[160] Motoyuki Oki, Koh Takeuchi, and Yukio Uematsu. 2018. Mobile network failure event detection and forecasting with multiple user activity data sets. In *AAAI 2018*.

[161] R González Perea, E Camacho Poyato, Pilar Montesinos, and JA Rodríguez Díaz. 2019. Prediction of irrigation event occurrence at farm level using optimal decision trees. *Computers and Electronics in Agriculture* 157 (2019), 173–180.
[162] Fotios Petropoulos and Spyros Makridakis. 2020. Forecasting the novel coronavirus COVID-19. *PloS One* 15, 3 (2020), e0231236.
[163] Lawrence Phillips, Chase Dowling, Kyle Shaffer, Nathan Hodas, and Svitlana Volkova. 2017. Using social media to predict the future: A systematic literature review. *arXiv preprint arXiv:1706.06134* (2017).
[164] Freddy Piraján, Andrey Fajardo, and Miguel Melgarejo. 2019. Towards a deep learning approach for urban crime forecasting. In *Workshop on Engineering Applications*. Springer, 179–189.
[165] R. J. Povinelli and Xin Feng. 2003. A new temporal pattern identification method for characterization and prediction of complex time series events. *TKDE* 15, 2 (2003), 339–352.
[166] Aimin Qi, Guosong Shao, and Wentong Zheng. 2018. Assessing China's cybersecurity law. *Computer Law & Security Review* 34, 6 (2018), 1342–1354.
[167] Zhi Qiao, Shiwan Zhao, Cao Xiao, Xiang Li, Yong Qin, and Fei Wang. 2018. Pairwise-ranking based collaborative recurrent neural networks for clinical event prediction. In *IJCAI 2018*. 3520–3526.
[168] Kira Radinsky, Sagie Davidovich, and Shaul Markovitch. 2012. Learning causality for news events prediction. In *WWW 2012*. 909–918.
[169] Kira Radinsky, Sagie Davidovich, and Shaul Markovitch. 2012. Learning to predict from textual data. *Journal of Artificial Intelligence Research* 45 (2012), 641–684.
[170] Kira Radinsky and Eric Horvitz. 2013. Mining the web to predict future events. In *WSDM 2013*. 255–264.
[171] Naren Ramakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self, et al. 2014. 'Beating the news' with EMBERS: forecasting civil unrest using open source indicators. In *SIGKDD 2014*. 1799–1808.
[172] Jonathan Rebane, Isak Karlsson, and Panagiotis Papapetrou. 2019. An investigation of interpretable deep learning for adverse drug event prediction. In *International Symposium on Computer-Based Medical Systems*. 337–342.
[173] D. Reid, A. Jaafar Hussain, H. Tawfik, R. Ghazali, and D. Al-Jumeily. 2018. Forecasting natural events using axonal delay. In *2018 IEEE Congress on Evolutionary Computation (CEC)*. 1–6.
[174] Honglei Ren, You Song, Jingwen Wang, Yucheng Hu, and Jinzhi Lei. 2018. A deep learning approach to the citywide traffic accident risk prediction. In *International Conference on Intelligent Transportation Systems*. 3346–3351.
[175] Jorge Reyes, Antonio Morales-Esteban, and Francisco Martínez-Álvarez. 2013. Neural networks to predict earthquakes in Chile. *Applied Soft Computing* 13, 2 (2013), 1314–1328.
[176] Alina Ristea, Mohammad Al Boni, Bernd Resch, Matthew S Gerber, and Michael Leitner. 2020. Spatial crime distribution and prediction for sporting events using social media. *IJGIS* (2020), 1–32.
[177] Mohammad Rostami, David Huber, and Tsai-Ching Lu. 2018. A crowdsourcing triage algorithm for geopolitical event forecasting. In *RecSys 2018*. 377–381.
[178] Bertrand Rouet-Leduc, Claudia Hulbert, Nicholas Lubbers, Kipton Barros, Colin J Humphreys, and Paul A Johnson. 2017. Machine learning predicts laboratory earthquakes. *Geophysical Research Letters* 44, 18 (2017), 9276–9282.
[179] Cynthia Rudin, Rebecca J Passonneau, Axinia Radeva, Haimonti Dutta, Steve Ierome, and Delfina Isaac. 2010. A process for predicting manhole events in Manhattan. *Machine Learning* 80, 1 (2010), 1–31.
[180] Shakila Khan Rumi, Ke Deng, and Flora D Salim. 2018. Theft prediction with individual risk factor of visitors. In *SIGSPATIAL 2018*. 552–555.
[181] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *WWW 2010*. 851–860.
[182] Felix Salfner, Maren Lenk, and Miroslaw Malek. 2010. A survey of online failure prediction methods. *ACM Computing Surveys (CSUR)* 42, 3 (2010), 1–42.
[183] Felix Salfner and Miroslaw Malek. 2007. Using hidden semi-Markov models for effective online failure prediction. In *2007 26th IEEE International Symposium on Reliable Distributed Systems (SRDS 2007)*. IEEE, 161–174.
[184] Jose A Sanchez, Aintzane Conde, Ander Arriandiaga, Jun Wang, and Soraya Plaza. 2018. Unexpected event prediction in wire electrical discharge machining using deep learning techniques. *Materials* 11, 7 (2018), 1100.
[185] Sara Santiso, Arantza Casillas, Alicia Pérez, Maite Oronoz, and Koldo Gojenola. 2014. Adverse drug event prediction combining shallow analysis and machine learning. In *Proceedings of the 5th International Workshop on Health Text Mining and Information Analysis (Louhi)*. 85–89.
[186] Jeffrey Shaman and Alicia Karspeck. 2012. Forecasting seasonal outbreaks of influenza. *PNAS* 109, 50 (2012), 20425–20430.
[187] Minglai Shao, Jianxin Li, Feng Chen, Hongyi Huang, Shuai Zhang, and Xunxun Chen. 2017. An efficient approach to event detection and forecasting in dynamic multivariate social media networks. In *WWW 2017*. 1631–1639.
[188] Yun Shen, Enrico Mariconti, Pierre Antoine Vervier, and Gianluca Stringhini. 2018. Tiresias: Predicting security events through deep learning. In *SIGSAC 2018*. 592–605.

[189] Seung Yeop Shin and Han-joon Kim. 2019. Autoencoder-based One-class Classification Technique for Event Prediction. In *Proceedings of the 2019 4th International Conference on Cloud Computing and Internet of Things*. 54–58.
[190] Prajwal Shrestha, Byung Suk Lee, and James P Bagrow. 2017. Predicting an effect event from a new cause event using a semantic web based abstraction tree of past cause-effect event pairs.. In *SIMBig*. 198–207.
[191] Aleksandr Simma and Michael I Jordan. 2012. Modeling events with cascades of Poisson processes. *arXiv preprint arXiv:1203.3516* (2012).
[192] Hagen Soltau, Hank Liao, and Haşim Sak. 2017. Neural speech recognizer: acoustic-to-word LSTM model for large vocabulary speech recognition. *Proc. Interspeech 2017* (2017), 3707–3711.
[193] Michael Spagat, Neil F Johnson, and Stijn van Weezel. 2018. Fundamental patterns and predictions of event size distributions in modern wars and terrorist campaigns. *PloS One* 13, 10 (2018), e0204639.
[194] Narayan Srinivasa, Qin Jiang, and Leandro G Barajas. 2008. High-impact event prediction by temporal data mining through genetic algorithms. In *International Conference on Natural Computation*, Vol. 1. 614–620.
[195] Zichun Su and Jialin Jiang. 2020. Hierarchical gated recurrent unit with semantic attention for event prediction. *Future Internet* 12, 2 (2020), 39.
[196] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: A core of semantic knowledge. In *WWW 2007*. 697–706.
[197] Gian Antonio Susto, Andrea Schirru, Simone Pampuri, Seán McLoone, et al. 2014. Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics* 11, 3 (2014), 812–820.
[198] Bayu Adhi Tama and Marco Comuzzi. 2019. An empirical comparison of classification techniques for next event prediction using business process event logs. *Expert Systems with Applications* 129 (2019), 233–245.
[199] James W Taylor. 2017. Probabilistic forecasting of wind power ramp events using autoregressive logit models. *European Journal of Operational Research* 259, 2 (2017), 703–712.
[200] Yue Teng, Dehua Bi, Guigang Xie, Yuan Jin, Yong Huang, Baihan Lin, Xiaoping An, Dan Feng, and Yigang Tong. 2017. Dynamic forecasting of Zika epidemics using Google trends. *PloS one* 12, 1 (2017), e0165085.
[201] Hannah Tops, Antal Van Den Bosch, and Florian Kunneman. 2013. Predicting time-to-event from Twitter messages. In *25th Benelux Conference on Artificial Intelligence, BNAIC 2013*. 207–214.
[202] Jay Ulfelder. 2013. A multimodel ensemble to forecast onsets of state-sponsored mass killing. In *American Political Science Association Annual Meeting. Available from: http://www. ssrn. com/abstract*, Vol. 2303048.
[203] Amin Vahedian, Xun Zhou, Ling Tong, Yanhua Li, and Jun Luo. 2017. Forecasting gathering events through continuous destination prediction on big trajectory data. In *SIGSPATIAL 2017*. 1–10.
[204] Amin Vahedian, Xun Zhou, Ling Tong, W Nick Street, and Yanhua Li. 2019. Predicting urban dispersal events: A two-stage framework through deep survival analysis on mobility data. In *AAAI 2019*, Vol. 33. 5199–5206.
[205] Kalyanaraman Vaidyanathan and Kishor S Trivedi. 1999. A measurement-based model for estimation of resource exhaustion in operational software systems. In *International Symposium on Software Reliability Engineering*. 84–93.
[206] Ricardo Vilalta and Sheng Ma. 2002. Predicting rare events in temporal domains. In *ICDM 2002*. IEEE, 474–481.
[207] Paul Voigt and Axel Von dem Bussche. 2017. The EU General Data Protection Regulation (GDPR). *A Practical Guide, 1st Ed., Cham: Springer International Publishing* (2017).
[208] Bao Wang, Xiyang Luo, Fangbo Zhang, Baichuan Yuan, Andrea L Bertozzi, and PJ Brantingham. 2018. Graph-based deep modeling and real time forecasting of sparse spatio-temporal data. *MiLeTSâĂŹ18, London, United Kingdom* (2018).
[209] Bao Wang, Penghang Yin, Andrea Louise Bertozzi, P Jeffrey Brantingham, et al. 2019. Deep learning for real-time crime forecasting and its ternarization. *Chinese Annals of Mathematics, Series B* 40, 6 (2019), 949–966.
[210] Chen Wang, Hoang Tam Vo, and Peng Ni. 2015. An IoT application for fault diagnosis and prediction. In *2015 IEEE International Conference on Data Science and Data Intensive Systems*. IEEE, 726–731.
[211] Dawei Wang and Wei Ding. 2015. A hierarchical pattern learning framework for forecasting extreme weather events. In *ICDM 2015*. IEEE, 1021–1026.
[212] Dawei Wang, Wei Ding, Kui Yu, Xindong Wu, et al. 2013. Towards long-lead forecasting of extreme flood events: a data mining framework for precipitation cluster precursors identification. In *SIGKDD 2013*. 1285–1293.
[213] Junxiang Wang, Yuyang Gao, Andreas Züfle, Jingyuan Yang, and Liang Zhao. 2018. Incomplete label uncertainty estimation for petition victory prediction with dynamic features. In *ICDM 2018*. IEEE, 537–546.
[214] Mingjun Wang and Matthew S Gerber. 2015. Using Twitter for next-place prediction, with an application to crime prediction. In *2015 IEEE Symposium Series on Computational Intelligence*. IEEE, 941–948.
[215] Qi Wang, Guangyin Jin, Xia Zhao, Yanghe Feng, and Jincai Huang. 2020. CSAN: A neural network benchmark model for crime forecasting in spatio-temporal scale. *Knowledge-Based Systems* 189 (2020), 105120.
[216] Qianru Wang, Junbo Zhang, et al. 2019. CityGuard: citywide fire risk forecasting using a machine learning approach. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–21.
[217] Zhongqing Wang and Yue Zhang. 2017. DDoS Event Forecasting using Twitter Data.. In *IJCAI 2017*. 4151–4157.

[218] Michael D Ward, Brian D Greenhill, and Kristin M Bakke. 2010. The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research* 47, 4 (2010), 363–375.
[219] Jeremy C Weiss and David Page. 2013. Forest-based point process for event prediction from electronic health records. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 547–562.
[220] Chuanxiu Xiong, Ajitesh Srivastava, Rajgopal Kannan, Omkar Damle, et al. 2019. On predicting crime with heterogeneous spatial patterns: methods and evaluation. In *SIGSPATIAL 2019*. 43–51.
[221] Cong Xue, Zehua Zeng, Yuanye He, Lei Wang, and Neng Gao. 2018. A MIML-LSTM neural network for integrated fine-grained event forecasting. In *International Conference on Big Data Technologies*. 44–51.
[222] Kazuo Yamaguchi. 1991. *Event history analysis*. Sage.
[223] Guolei Yang, Ying Cai, and Chandan K Reddy. 2018. Spatio-temporal check-in time prediction with recurrent neural network based survival analysis. In *IJCAI 2018*. 2976–2983.
[224] Jaewon Yang, Julian McAuley, Jure Leskovec, Paea LePendu, and Nigam Shah. 2014. Finding progression stages in time-evolving event sequences. In *WWW 2014*. 783–794.
[225] Qiang Yang, Hui Wang, and Wei Zhang. 2002. Web-log mining for quantitative temporal-event prediction. *IEEE Computational Intelligence Bulletin* 1, 1 (2002), 10–18.
[226] Yiying Yang, Zhongyu Wei, Qin Chen, and Libo Wu. 2019. Using External Knowledge for Financial Event Prediction Based on Graph Neural Networks. In *CIKM 2019*. 2161–2164.
[227] Fei Yi, Zhiwen Yu, Fuzhen Zhuang, and Bin Guo. 2019. Neural network based continuous conditional random field for fine-grained crime prediction.. In *IJCAI 2019*. 4157–4163.
[228] Fei Yi, Zhiwen Yu, Fuzhen Zhuang, Xiao Zhang, and Hui Xiong. 2018. An integrated model for crime prediction using temporal and spatial factors. In *ICDM 2018*. IEEE, 1386–1391.
[229] James Yonamine. 2013. Predicting future levels of violence in afghanistan districts using Gdelt. *Unpublished manuscript* (2013).
[230] Kui Yu, Dawei Wang, Wei Ding, Jian Pei, David L Small, Shafiqul Islam, and Xindong Wu. 2015. Tornado forecasting with multiple Markov boundaries. In *SIGKDD 2015*. 2237–2246.
[231] Shuqi Yu, Linmei Hu, and Bin Wu. 2019. DRAM: A deep reinforced intra-attentive model for event prediction. In *International Conference on Knowledge Science, Engineering and Management*. Springer, 701–713.
[232] Sheng Yu and Subhash Kak. 2012. A survey of prediction using social media. *arXiv preprint arXiv:1203.1647* (2012).
[233] Zhuoning Yuan, Xun Zhou, and Tianbao Yang. 2018. Hetero-convLSTM: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In *SIGKDD 2018*. 984–992.
[234] Weiran Yuchi, Jiayun Yao, Kathleen E. McLean, et al. 2016. Blending forest fire smoke forecasts with observed data can improve their utility for public health applications. *Atmospheric Environment* 145 (2016), 308 – 317. https://doi.org/10.1016/j.atmosenv.2016.09.049
[235] Jenny Yuen and Antonio Torralba. 2010. A data-driven approach for event prediction. In *ECCV 2010*. Springer, 707–720.
[236] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. 2014. *Social media mining: An introduction*. Cambridge University Press.
[237] Qian Zhang, Nicola Perra, Daniela Perrotta, Michele Tizzoni, Daniela Paolotti, and Alessandro Vespignani. 2017. Forecasting seasonal influenza fusing digital indicators and a mechanistic disease model. In *WWW 2017*. 311–319.
[238] Liang Zhao, Feng Chen, Jing Dai, Ting Hua, Chang-Tien Lu, and Naren Ramakrishnan. 2014. Unsupervised spatial event detection in targeted domains with applications to civil unrest modeling. *PloS one* 9, 10 (2014), e110206.
[239] Liang Zhao, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2015. Spatiotemporal event forecasting in social media. In *SDM 2015*. SIAM, 963–971.
[240] Liang Zhao, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2016. Multi-resolution spatial event forecasting in social media. In *ICDM 2016*. IEEE, 689–698.
[241] Liang Zhao, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2016. Online spatial event forecasting in microblogs. *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 2, 4 (2016), 1–39.
[242] Liang Zhao, Jiangzhuo Chen, Feng Chen, Wei Wang, Chang-Tien Lu, and Naren Ramakrishnan. 2015. Simnest: Social media nested epidemic simulation via online semi-supervised deep learning. In *ICDM 2015*. IEEE, 639–648.
[243] Liang Zhao, Olga Gkountouna, and Dieter Pfoser. 2019. Spatial auto-regressive dependency interpretable Learning Based on Spatial Topological Constraints. *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 5, 3 (2019), 1–28.
[244] Liang Zhao, Qian Sun, Jieping Ye, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2015. Multi-task learning for spatio-temporal event forecasting. In *SIGKDD 2015*. 1503–1512.
[245] Liang Zhao, Qian Sun, Jieping Ye, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2017. Feature constrained multi-task learning models for spatiotemporal event forecasting. *TKDE* 29, 5 (2017), 1059–1072.

[246] Liang Zhao, Junxiang Wang, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2017. Spatial event forecasting in social media with geographically hierarchical regularization. *Proc. IEEE* 105, 10 (2017), 1953–1970.

[247] Liang Zhao, Junxiang Wang, and Xiaojie Guo. 2018. Distant-supervision of heterogeneous multitask learning for social event forecasting with multilingual indicators. In *AAAI 2018*.

[248] Liang Zhao, Jieping Ye, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2016. Hierarchical incomplete multi-source feature learning for spatiotemporal event forecasting. In *SIGKDD 2016*. 2085–2094.

[249] Sendong Zhao, Quan Wang, Sean Massung, Bing Qin, Ting Liu, Bin Wang, and ChengXiang Zhai. 2017. Constructing and embedding abstract event causality networks from text snippets. In *WSDM 2017*. 335–344.

[250] Xiangyu Zhao and Jiliang Tang. 2017. Modeling temporal-spatial correlations for crime prediction. In *CIKM 2017*. 497–506.

[251] Qiuzhen Zhong, Jingjing Wang, Xuejie Meng, Siqing Liu, and Jiancun Gong. 2019. Prediction model for solar energetic proton events: Analysis and verification. *Space Weather* 17, 5 (2019), 709–726.

[252] Cheng Zhou, Boris Cule, and Bart Goethals. 2015. A pattern based predictor for event streams. *Expert Systems with Applications* 42, 23 (2015), 9294–9306.

[253] Lihua Zhou, Guowang Du, Ruxin Wang, Dapeng Tao, Lizhen Wang, Jun Cheng, and Jing Wang. 2019. A tensor framework for geosensor data forecasting of significant societal events. *Pattern Recognition* 88 (2019), 27–37.