

数据驱动的事件预测技术最新研究进展

李志鹏^{1,2}, 杨阳朝², 廖勇^{1,2}, 俞能海¹, 吴哲²,
谢海永¹, 石珺², 曾曦²

¹ 中国科学技术大学网络空间安全学院 合肥 中国 230027

² 深圳市网联安瑞网络科技有限公司 深圳 中国 518038

摘要 “事件”(event)是指在特定时空发生的对人类社会和自然界产生较为明显影响的事情。社会动乱、暴恐事件、传染病大流行等例子是给国家和社会安全带来严重威胁的“事件”。如果能够提前对这些事件的发生进行有效预测,将有助于做好应对准备,大大减少不必要的损失,因此事件预测技术在实际中具有重大社会应用价值,能够在社会安全、风险感知、传染病防控等方面发挥重要作用。对事件进行科学准确的预测曾经是一个十分具有挑战性的问题,近期大数据和数据挖掘的发展为事件预测技术带来了新的机遇。本文就以数据驱动的事件预测技术最新研究进展做一系统化的综述,介绍了事件预测的形式化建模与性能度量指标,对事件预测技术领域的最新研究成果进行了科学归类与总结,分为频繁模式挖掘、传统分类模型、时间序列预测、时序点过程、地理空间位置预测、事件图谱、无监督方法、多技术融合预测八大类方法,将每类方法做了系统地阐述,接着探讨了事件预测技术的主要应用领域,最后展望了这一技术未来面临的挑战和潜在的研究方向,以期进一步推动事件预测技术的发展与应用。

关键词 计算社会学; 机器学习; 数据挖掘; 事件图谱; 事件预测

中图分类号 TP309.2 DOI号 10.19363/J.cnki.cn10-1380/tn.2022.01.03

A Survey of Recent Advances in Data-Driven Event Prediction Research

LI Zhipeng^{1,2}, YANG Yangzhao², LIAO Yong^{1,2}, YU Nenghai¹, WU Zhe², XIE Haiyong¹,
SHI Jun², ZENG Xi²

¹ School of Cyberspace Science, University of Science and Technology of China, Hefei 230027, China

² Shenzhen CyberAray Co., Ltd., Shenzhen 518038, China

Abstract An event is usually defined as an incident that takes place in a specific location and time, impacting either the society or the nature in a nontrivial way. Civil unrest, terrorist attacks and pandemics are commonly known events that can pose serious threats to both national security and public safety. Effectively predicting upcoming events is highly desirable in reality, as successful predictions can be of great use in designing countermeasures to prevent or reduce potential losses. Predicting events in advances plays an important role in public security, risk perception and infectious disease prevention and control, etc. Although event prediction used to be a technically challenging task, recent advances in the fields of big data analytics and machine learning have brought promising opportunities in applying these techniques to solve real world prediction problems. This paper presents a systematic survey on data-driven event prediction research studies. We first introduce the formal definition of event prediction problem and the evaluation metrics of various prediction techniques. Then, the state-of-the-art algorithms and schemes proposed in the field of event prediction are summarized and classified. All existing event prediction methods can be classified into 8 categories: frequent pattern mining, traditional classification model, time series model, temporal point processes, geospatial predictive modeling, event knowledge graph, unsupervised machine learning, multi-model fusion method. We present a systematic and comprehensive summary for each category of these methods. After that, we proceed by discussing the main real-world applications of event prediction techniques, including public security, disease prevention, smart city and natural disaster prediction. We conclude this survey by summarizing multiple open research problems and several possible directions on event prediction research. To the best of our knowledge, this study provides a comprehensive summarization of recent research and advances on data-driven event prediction. We hope this paper can be a useful booster in the research of event prediction and applying this technique in solving real world problems.

Key words computational social science; machine learning; data mining; event graph; event prediction

通讯作者: 俞能海, 博士, 教授, Email: ynh@ustc.edu.cn.

本课题得到国家自然科学基金联合基金项目(No. U19B2044)资助。

收稿日期: 2020-12-19; 修改日期: 2021-04-27; 定稿日期: 2021-11-10

1 引言

计算机科学是研究最活跃、发展最迅速、影响最广泛的学科之一,已成为人类解决全球性挑战的重要技术途径。个人计算机、互联网、大数据、人工智能、量子计算,计算机科学的发展正在深刻地影响着人类社会的发展进程。现阶段的科学研究正处于一个知识大融合、大交叉的时代,在技术推动和需求牵引双重作用下,计算机科学也在不断催化新型交叉学科的融合发展。计算机科学为社会科学研究提供了新的途径和方法,2009年哈佛大学教授 Lazer 等 15 名国际顶尖学者在 *Science* 杂志合作发表论文^[1],正式提出计算社会科学(Computational Social Science)这一概念,标志着通过大规模采集和分析社会学数据进而揭示个人和群体行为模式这一新学科领域的形成。研究者们开始尝试广泛收集和利用互联网大数据(如新闻、论坛、博客、社交网络等)来感知、甚至预知社会,实现社会现象或社会事件的发现感知与科学预测。

人类进行的社会活动往往是由事件驱动的,社会事件反映了短期或瞬时的人类活动及相应影响。事件是人类社会的核心特征之一,是对现实社会变化的原子性描述,事件演化的过程可以看成是现实世界和人类活动的抽象映射。美国国防高级研究计划局 DARPA 资助的话题检测与跟踪项目(Topic Detection and Tracking, TDT)最早针对新闻媒体信息流中新事件的自动检测与跟踪进行研究,美国国家标准技术研究所 NIST 组织主办的 ACE 测评会议从 2005 年起将事件检测和事件抽取作为基本任务之一,自然语言处理领域的研究人员对此进行了大量研究^[2-3]。基于自然语言处理和分布式计算等领域的研究成果,自动化构建大规模结构化事件库的技术逐渐成熟。Google Jigsaw 赞助的 GDELT(Global Database of Events, Language, and Tone)项目^[4]利用超过 100 种语言的全球新闻媒体数据,自动发现并记录了自 1979 年 1 月 1 日以来的所有人类社会主要事件,掀起了计算社会学的研究热潮,仅通过 Google Scholar 可检索到的基于 GDELT 事件库进行研究的文献就有 1700 多篇,这一方面展现了学者对于利用海量数据研究社会事件这一方向的高度兴趣,另一方面也显示这一研究领域所具有的实际价值。

相比于事件检测和事件抽取任务,事件预测主要关注未来预期将会发生的事件。事件预测技术通过对相关数据的深入挖掘与分析,运用科学的知识、

手段和方法,对事件的状态或未来发展趋势做出科学的估计和评价。在社会安全、传染病预防、智慧城市、自然灾害预测等应用领域,如果能够提前预知事件的发生状态或未来走势,提前做出判断,就可以对事件进行预测和预警,有利于适时调整计划,采取措施实施精准调控,能够带来显著的经济收益,减少不必要的损失,具有重要的现实意义。准确的事件预测结果对于生产生活中的趋利避害、计划决策起着至关重要的作用。国内外的研究者近年来针对事件预测技术开展了大量研究,积累了丰富的事件预测方法,涌现了一大批优秀的学术成果。然而目前缺少对事件预测技术研究进展的全面性的总结和系统性的阐述,本文首次系统分析和总结了各类事件预测技术,对事件预测问题的形式化模型与性能度量进行了分析,并对事件预测技术的实际应用和研究展望进行了探讨。

本文的结构如下:第 2 节基于现有文献建立了事件预测的形式化模型,介绍了事件预测算法的性能度量指标;第 3 节事件预测技术进行了总结和归纳,介绍了事件预测的主要方法和最新研究进展;第 4 节介绍了事件预测技术在实际生产生活中的应用;第 5 节对事件预测技术的未来研究进行了展望;第 6 节是结束语,对全文进行了总结。

2 事件预测问题建模与性能度量

人们可能会对“事件到底是不是可预测的”即事件可预测性这个问题有所疑问,已有大量文献通过真实实验数据表明了使用数据分析的方法对事件进行预测的技术可行性与有效性,Muthiah 等人^[5]通过 4 年多社会内乱(Civil Unrest)事件预测的实际研发部署与运维经验,在数据挖掘顶级学术会议 KDD 上发表文章对限制事件预测技术可预测性和效果的两大类不确定性因素(Uncertainty)做了非常好的叙述,分别为起因(Cause)和时机(Timing),如图 1 所示,本文对此总结如下。

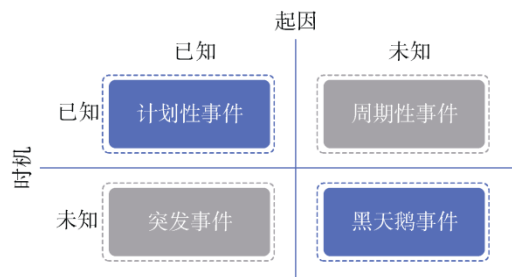


图 1 事件预测不确定因素

Figure 1 The uncertainties of event prediction

如果知道事件发生的起因和时机, 例如社会团体通过海报宣传等方式组织游行事件, 游行的时间和地点都有事先通知预告, 这类事件称之为计划性事件(Planned Events)。Muthiah 等人^[6]发现此类事件占社会内乱事件的绝大多数, 尤其大规模的游行事件往往需要提前宣传和组织公告, 在互联网社交媒体等渠道发布时间和地点信息, 以便形成共识。此类事件预测问题已经有很好的解决方法, 详见本文 3.8 节中关于计划性事件预测的介绍。

周期性事件(Recurring Events)经常在每年特定的时间点发生, 但事件发生的原因和参与者的诉求可能多种多样, 例如在具有较大数量穆斯林人口的国家, 抗议和暴力事件经常发生在周五集体礼拜“主麻日”之后, 因为此时人群较为聚集。这类事件的预测问题可以通过本文 3.1 节频繁模式挖掘、3.3 节时间序列模型、3.4 节中时空关联预测等多种方法解决。

事件爆发的起因已知, 但是事件发生的时间具有偶然性的事件称为突发事件(Spontaneous Events)。例如“阿拉伯之春”系列事件, 事件的起因包括经济衰退、失业率高、政府贪污腐败等社会经济因素。此类事件的预测往往可以通过广泛收集开源的互联网大数据信息, 结合本文中总结的多类事件预测技术进行研究, 受到计算社会学的广泛关注。针对这类事件爆发的导火索进行预测或检测, 具有极高的应用价值, 近年来成为数据挖掘领域的研究热点。

最后, 黑天鹅事件(Black Swan Events)发生的时间和起因都是未知的。此类事件一般是离群值(Outlier), 因为它出现在人们的期望范围之外, 基于过去历史经验往往会得出其不可能出现的结论。黑天鹅事件的发生可能会产生极端影响, 一般被认为不可预测, 但是黑天鹅事件的检测可以通过第 1 节引言中介绍的事件检测和事件抽取技术实现, 黑天鹅事件的后续演化预测可以利用本文中介绍的事件预测技术。

2.1 形式化模型

参考 DARPA 资助的 TDT 项目, 将事件定义为在特定时间和特定地点发生的某件事情“Event is something that happens at a particular time and place”^[7]。因此本文将事件形式化定义为三元组

$$e = (t, l, d) \in \mathcal{E},$$

其中, $t \in \mathcal{T}$ 表示事件发生的时间, $l \in \mathcal{L}$ 表示事件发生的地点, $d \in \mathcal{D}$ 表示事件发生的细节详情(Details)。 \mathcal{T} , \mathcal{L} , \mathcal{D} 构成了事件预测问题空间的时间域、地点域和详情域, 即

$$\mathcal{E} = \mathcal{T} \times \mathcal{L} \times \mathcal{D}$$

例如, (2012 年 8 月 9 日, “巴西, 北里奥格兰德”, “抗议当地市政厅, 要求减少车费价格”)就是一个事件。

事件预测问题可能还会有额外的信息作为输入, 一般会将其进行量化, 本文统称为特征(Feature), 形式化定义为 $f \in \mathcal{F}$ 。将当前时间记为 t_0 , 事件预测问题可以形式化建模为: 输入历史事件及补充特征 $X \in E \times F, t \leq t_0$, 输出 $\hat{e} \in \hat{E}, t \geq t_0$, 即未来事件发生的时间、地点或详情, 优化目标是输出的预测 \hat{e} 能够符合实际情况, 在预测的时间点得到验证。值得指出的是, 实际事件预测问题可能是该形式化模型的子集, 比如人们可能只关心事件发生的时间。

2.2 性能度量指标

本小节对事件预测相关技术的性能评估方法和指标做简要的总结。2.1 节提出事件预测算法的目标是输出的预测符合实际发生事件的检验, 这需要有衡量“符合”这一文字描述的评价标准, 本小节对此进行量化。由于事件是三元组 $e = (t, l, d)$ 形式, 不能一概使用机器学习和统计学领域常用的准确率(Precision)、召回率(Recall)、 F 值等指标, 需要根据实际问题情况确定。

除了准确率、召回率、 F 值这类标准指标外, 对于只关注事件地点的情况, 还可以使用预测地点与实际地点的欧氏距离、曼哈顿距离等来衡量, 如果是非欧几何空间如网络中的节点, 可以使用预测点与真实点的最短路径来衡量。对于事件时间, 可以使用预测发生时刻与实际发生时刻的绝对值、均方差等来衡量, 预测模型输出预测结果的时刻与预测发生时刻的差值即提前预警时间也是一个重要指标。对于事件详情, 如果是事件规模如 0~5 分级表示, 可采用分类模型的经典度量指标, 如果是事件的情况描述, 可以采用 NLP 领域中的文本度量标准, 例如编辑距离(Edit Distance)、BLEU 等。对于关注事件预测结果 2 个及多个组合的情况, 可以定义相应的加权组合值作为最后的评价标准。

3 事件预测技术

使用“Event Prediction”和“Event Forecasting”作为关键词, 在 IEEE Xplore、ACM Digital Library、Elsevier ScienceDirect、Springer Link、Google Scholar 等学术文献数据库和学术文献搜索引擎进行检索, 从中挑选出了相关高水平会议和期刊文章 100 余篇,

大部分为 2015 年至今发表的文献(75%+), 对其中具有代表性的事件预测技术进行了总结和归纳, 将现有文献中的事件预测技术划分为 8 大类, 分别为频繁模式挖掘、传统分类模型、时间序列预测、时序点过程、地理空间位置预测、事件图谱、无监督方法、多技术融合预测。由于对检索到的文献进行了精挑细选, 本文或许没有覆盖所有事件预测相关的文献, 但是我们相信本文已经较为全面地给出了事件预测技术总体情况(Big Picture)和最新的(State-of-the-art)研究进展。

本文中事件预测技术的划分标准参考了数据挖掘领域中从不同角度对数据进行分析的方法, 同时结合了事件预测技术当前研究热点的独特性。数据挖掘方法主要包括频繁模式挖掘(关联分析等)、传统分类模型(朴素贝叶斯、决策树、支持向量机、K 近邻等)、无监督方法(聚类、异常检测等)^[8]。基于文献检索结果, 存在较多研究将统计、金融领域的时间序列分析, 随机过程中的点过程方法用于事件预测。伴随知识图谱成为人工智能领域的研究热点, 事件知识图谱也成为事件预测领域的重要研究方向。事件预测问题有时更为关注地理空间方面的因素, 因此本文将地理空间位置预测方法单独列出。多技术融合预测主要关注将多种事件预测方法融合, 以及业界中具有较大影响力的事件预测实际系统。

3.1 频繁模式挖掘

此类方法尝试从历史事件序列数据中挖掘出特定的频繁模式, 从而进行特定事件的预测。Vilalta 等人^[9]提出了一种指定类型事件预测的算法, 该算法将事件预测问题转化为搜索目标类别事件发生前的频繁事件集。Srivatsan 等人^[10]设计了一种事件频繁片段(Frequent episode)挖掘框架, 片段定义为一小段连续的事件序列, 频繁片段发现就是挖掘事件流中的时序模式。在此基础上, Laxman 等人^[11]将该框架扩展为支持多个事件序列的数据集。Zhou 等人^[12]进一步对预测结果按照置信度(Confidence)进行排序, 并且输出 top-k 事件。总的来说, 基于频繁模式挖掘的事件预测技术可以解决事件发生的时间间隔不恒定、多种类别事件预测、事件样本类别不平衡的问题, 但只能预测少量具备特定模式的事件, 更为成熟准确的方法会将频繁模式挖掘的规则与其他方法如分类模型进行结合^[13]。

3.2 传统分类模型

这类技术的主要途径是将事件预测任务通过一定的转换, 建模为经典的机器学习分类问题, 例如预测事件是否会发生可以转化为简单二分类模型:

事件发生(正类)、事件未发生(负类)。Ballings 等人^[14]采用 logistic 回归、分类树模型建立客户流失事件预测模型, Caigny 等人^[15]融合了 logistic 回归和决策树模型, 提出一种新颖的 logit leaf 模型对客户流失事件进行预测, 在精度(Accuracy)和模型可解释性上都达到了很好的效果。Kallus^[16]通过采集 30 多万个涵盖 7 种语言的全球 Web 网站内容, 基于随机森林模型, 对群体行为(Crowd behavior)进行预测。Korkmaz 等人^[17]融合 Twitter、博客、Tor 流量日志、GDELT、货币汇率等多种数据源, 提取相应特征进行 Logistic 回归分类, 来预测社会内乱(Civil unrest)事件是否会发生。Zhao 等人^[18]基于 LASSO 回归, 对给定的地理位置在下一个时间点是否会发生事件进行预测, 提出了一种有效的参数优化求解算法保证得到最优值, 在南美洲巴西、委内瑞拉等国社会内乱事件实际数据集上取得了较好的测试性能。Santiso 等人^[19]使用电子病历(Electronic health records)数据, 学习训练随机森林分类器模型, 对是否会药物不良反应事件(Adverse drug reactions)进行预测。

使用经典机器学习算法解决事件是否发生的二分类问题对训练数据有一定的要求。以预测某建筑是否会发生火灾事件为例, 对分类算法进行训练时, 需要包含正类的发生火灾数据和负类的未发生火灾数据, 然而实际数据中可能只包含发生火灾的建筑这一类数据, 并不包含未发生火灾的建筑数据。针对这一问题, Shin 等人^[20]提出了一种基于自编码(AutoEncoder)的单分类(One-class Classification)预测技术, 能够有效地预测有火灾事件风险的建筑。

针对事件多分类问题, Tops 等人^[21]利用 Twitter 数据对事件发生时间进行预测, 通过将时间自动划分为多个离散时间段, 建模为多分类问题, 并对朴素贝叶斯(Naïve Bayes)、k 近邻(k-Nearest Neighbors, k-NN)、支持向量机(Support Vector Machines, SVM)模型进行了事件发生时间的多分类预测和性能对比。事件分级常用于使媒体和公众更直观了解事件的严重性和紧急程度, 如红色、橙色、黄色、蓝色预警等级划分, Gao 等人^[22]基于有序分类(Ordinal classification)对事件级别进行了预测。Tama 等人^[23]对 5 大类 20 个传统分类模型在业务流程(Business Process)事件预测上的性能表现进行了测试。基于分类模型进行事件预测是较为普遍和朴素的想法, 也是被研究得最广泛、最为经典的方法^[24]。

3.3 时间序列预测

时间序列是一组按照发生时间先后顺序进行排列的数据点序列, 通常一组时间序列内数据点的时

间间隔为恒定值, 例如 1 秒、1 天、1 周、1 年等, 时间序列分析方法在经济学、金融学、信号处理等学科领域中有广泛应用, 近年来在事件预测技术当中也得到了较多研究和关注。时间序列预测可以形式化描述为基于观察到的历史时间序列 $X = \{x_t, t = 1, \dots, N\}$, 预测未来时间序列 $\hat{X} = \{x_t, t = N + 1, \dots, T\}$, 例如 x_t 可以表示关注地每天发生的事件数量。传统时间序列预测主要采用线性模型, 自回归模型 (Auto-Regressive, AR) 假定 x_t 与一个或者多个之前的时间序列值为线性关系, 即 $x_t = c + \sum_{i=1}^p \varphi_i x_{t-i} + \varepsilon_t$, 含义为 x_t 等于 p 个 (p 可以为 1) 前值的线性组合, 加上常数项和随机误差。移动平均模型 (Moving Average model, MA) 使用历史预测误差来建立用于预测未来值的模型。自回归移动平均模型 (Auto-Regressive Moving Average, ARMA) 由自回归模型与移动平均模型为基础混合构成, 这三个模型都假定时间序列是平稳的 (stationary)。对于非平稳时间序列, 可以采用差分自回归移动平均模型 (Auto-Regressive Integrated Moving Average, ARIMA), 该模型对 ARMA 做了扩展, 对非平稳时间序列数据进行差分运算, 转化为平稳时间序列处理。分数阶自回归移动平均模型 (Auto-Regressive Fractionally Integrated Moving Average, AFIMA) 是基于分数阶差分的 ARMA 模型, 可以看作是 ARIMA 模型更通用化的形式扩展, Yonamine^[25] 使用 ARFIMA 模型对阿富汗地区发生暴乱事件的程度进行了预测分析。Provinelli 等人^[26] 提出了一种新的识别事件时间序列时序模式的方法, 通过采用基于遗传算法的优化方法搜索最优的时序模式, 能够有效地提取事件时间序列特征, 对事件进行预测。具体来说, 利用时延嵌入 (Time-delayed Embedding) 方法, 将时间序列 $X = \{x_t, t = 1, \dots, N\}$ 映射为 \mathcal{R}^Q 空间的列向量 $\hat{x}_t = (x_{t-(Q-1)\tau}, \dots, x_{t-\tau}, x_t)^T$, 其中 τ 为时延, 之后构造事件特征函数 $g(\hat{x}_t)$ 将嵌入向量 \hat{x}_t 映射为事件发生的可能性, 最后基于数据集的标签进行训练。

隐马尔可夫模型 (Hidden Markov Model, HMM) 用于时间序列预测也吸引了大量研究者的关注^[27]。与传统时间序列预测方法类似, HMM 这类方法假设事件的过去延续到将来, 从而可以根据过去的事件变化趋势进行外推和引申, 预测事件未来的发展趋势。Petroff 等人^[28] 将恐怖事件分成 3~5 个状态, 给定输入序列数据, 使用 HMM 搜索最有可能生成该序列数据的隐藏序列状态, 进而预测将来时间序列上的恐怖事件结果, 并且在阿富汗和伊拉克等实际数据集上进行了实验。Salfner 等人^[29] 将 HMM 用于通

信系统故障 (System Failure) 事件检测, 实验结果表明该方法非常有效。其他基于生成模型的方法还包括 Antunes 等人^[30] 提出的基于 Bayesian 方法的事件时间序列预测。

3.4 时序点过程

传统分类模型适用于预测事件是否会发生等可转化为分类问题的事件预测, 而时间序列模型一般假设序列内数据点的时间间隔等长, 这都导致了一定程度上的适用性问题。Minor 等人^[31] 对智慧家庭 (Smart Home) 场景下的活动事件预测 (Activity Prediction) 问题进行了研究, 基于回归树模型 (Regression Tree) 对事件将要发生的确切时间 (连续时间) 进行了预测, 但总的来说, 传统回归模型如线性回归基于正态分布的假设不太适用于现实生活中的事件预测。时间序列模型一般预先设定好序列内数据点的时间间隔, 将事件序列转换为多段等长的时间间隔, 进而聚合每段时间间隔内的事件, 形成离散事件序列点。基于等间隔采样形成的离散特性, 引入了事件时间序列聚合过程中的精确时间戳信息丢失问题。针对该问题, 时序点过程 (Temporal Point Process) 模型对连续时间域上的异步事件序列预测进行建模, 如图 2 所示, 可以实现对历史事件时间信息的充分利用, 这类模型已经成为一种重要的事件预测问题解决方法。

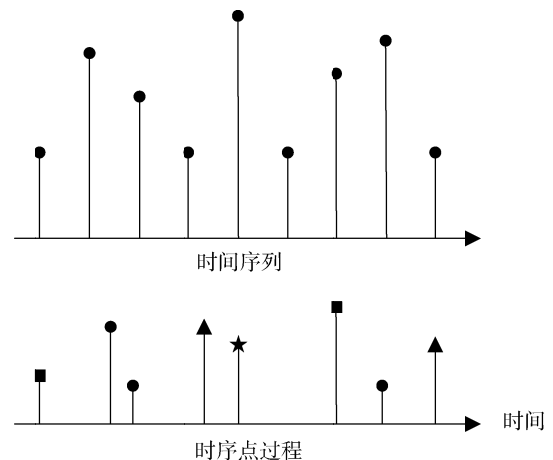


图 2 时间序列与时序点过程对比

Figure 2 Time series and temporal point process

时序点过程模型的核心是条件强度函数 (Conditional Intensity Function), 形式化定义如下:

$$\begin{aligned} \lambda(t|X_t) &= P(\text{events in } [t, t + dt] | X_t) \\ &= \mathbb{E} \left[\frac{N(t, t + dt)}{dt} \middle| X_t \right] = \frac{f(t|X_t)}{1 - F(t|X_t)} \end{aligned}$$

其数学含义为基于时刻 t 之前的历史事件 X_t 在无

穷小的时间窗口 $[t, t + dt)$ 内发生事件次数的条件期望。其中, $f(t|X_t)$ 是条件密度函数, $F(t|X_t)$ 是其对应的累积分布函数。进一步推导可以得到:

$$f(t|X_t) = \lambda(t|X_t) \cdot \exp\left(-\int_{t_0}^t \lambda(u|X_t)du\right)$$

一般通过极大似然估计(Maximum Likelihood Estimation)对上述模型进行训练, 预测发生事件的下一个时间点为:

$$\hat{t} = \int_{t_0}^{+\infty} t \cdot f(t|X_t)dt$$

传统时序点过程建模前需要人工预先确定相应的条件强度函数, 现有基于时序点过程的事件预测文献中, 采用的典型时序点过程模型包括 Poisson 过程^[32-33], Hawkes 过程^[34], Weibull 过程^[35]。传统时序点过程有多种函数形式, 其建模的成功往往依赖于正确的模型选择。如果人工预先选择的强度函数和事件数据的规律是不相符, 那模型的效果往往大打折扣。随着深度学习技术在多个领域展现出强大的能力, 基于深度点过程(也被称为神经点过程)进行事件预测^[36]逐渐成为一个新兴研究热点。利用神经网络的强大容量, 通过大规模数据挖掘隐藏的事件规律, 试图学习拟合能力更强的模型, 并减少对先验知识的依赖, 从而对未来事件进行预测, 但是一定程度上牺牲了模型可解释性。

目前大多数的时序点过程方法都假设事件历史数据是完整的序列, 但是这个条件在真实场景下难以保证。对于存在删失数据(Censored Data)、截断数据(Truncated Data)的情况, 有文献研究利用生存分析(Survival Analysis)框架^[37-38]处理这一问题。

3.5 地理空间位置预测

前述事件预测技术, 主要针对事件将要发生的时间, 或者未来时间点的事件属性例如规模、类型进行预测, 在很多实际应用中, 除了发生的时间, 我们还关注事件发生的地理空间位置, 此类方法主要研究如何对事件的地点进行预测^[39]。根据地理空间位置采用模型的不同, 可以分为基于地理空间网格(Grid)的方法和基于网络和节点的方法, 如图 3 所示。地理空间网格方法通过将一定规则组织起来的、连续的、多分辨率的网格单元, 逐步逼近最真实的地形, 把空间的不确定性因素控制在相应的尺度范围之内, 这也是测绘学、地理学最常用的方法。基于网络和节点的方法是对地理位置的另一种形式的抽象, 将地点抽象表示成网络或图数据结构, 可以类比现实生活中的网络连接图、地铁路线图等。实际上, 事件预测问题往往需要同时考虑空间和时间两个方面的

因素, 即时空预测问题, 这是时空数据挖掘中的一个重要任务。

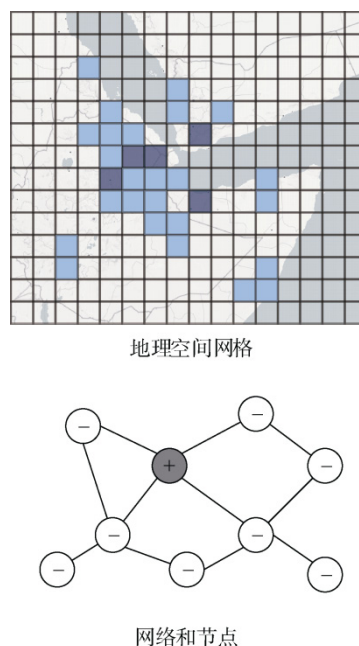


图 3 地理空间网格与网络和节点对比
Figure 3 Grid-based vs. network and node-based

Wang 等人^[40]设计了一种层次化的时空模式学习架构, 用于暴雨等极端天气事件预测, 提出了 SCOT 优化算法将第一层的输入进行空间聚类, 将网格转化为实际的预测区域。Xiong 等人^[41]发现对地理位置进行划分的方法极大影响犯罪事件预测的准确性, 因此提出了一种基于犯罪数量的启发式算法将地理位置划分为不同的区域, 进而使用分类模型对这些区域进行训练, 最后对各区域的犯罪事件数量进行预测。Hao 等人^[42]使用核密度估计(Kernel Density Estimation, KDE)方法, 将全球恐怖主义事件库(Global Terrorism Database, GTD)中的原始事件数据(点数据)转化为网格数据, 然后采用随机森林方法进行训练和预测。Ristea 等人^[43]分析了犯罪率、人口统计数据、社会经济因素、带有地理位置的推文等变量之间的空间关系, 对体育活动事件是否会有暴力犯罪活动进行了预测, 研究发现借助包含暴力信息的推文, 可以提高预测模型的准确性。Bao 等人^[44]提出了一种时空深度学习架构模型, 用于预测城市范围内的车辆碰撞事件风险, 并且在不同时空分辨率的条件下进行了对比实验, 该模型优于常用的计量经济学和机器学习模型。基于网格的地理信息数据跟基于像素的图片具有一定的相似性, Mukhina 等人^[45]基于 Instagram 社交媒体数据, 使用卷积神经网络(Convolutional Neural Network, CNN), 预测可能会

发生事件的地区。Wang 等人^[46]采用一种时空残差网络(ResNet)对洛杉矶的犯罪事件分布进行预测, 并且给出了一种三元数值优化方法解决实际使用部署过程中的计算资源消耗问题。Vahedian 等人^[47-48]通过对不完整运动轨迹的分析来预测将会发生的群体聚集型事件, 预测模型包括目的地预测和事件预测两部分, 通过目的地预测模型得到的概率分布, 对目标地点给定时间潜在到达人群数量进行预测, 进而预测潜在的聚集事件。

多任务学习(Multi-Task Learning, MTL)是机器学习中的一个很有前景的领域, 其目标是利用多个学习任务中所包含的有用信息来帮助每个任务得到更为准确的性能。Zhao 等人^[49]最先把 MTL 引入时空事件预测领域, 将对一个国家中多个城市同时进行事件预测建模为多任务学习问题, 综合使用 LASSO 模型的静态特征和 DQE 模型的动态特征, 开发了高效求解算法, 在墨西哥、巴西等 4 个拉丁美洲国家的实际数据集上测试取得了较好的预测效果, 在流感爆发数据集上测试也得到了有效性验证^[50]。Zhao 等人^[51]进一步将 MTL 模型扩展为支持地理空间层次结构(如国家、州、城市 3 级), 基于 ADMM 和动态规划算法进行参数求解。Ning 等人^[52]提出的多任务学习框架还支持前导事件(Precursor)的发现, 所谓前导事件, 指发生在给定事件之前, 并对给定事件有关联或者有影响的事件。

研究者们还探索了其他时空预测模型的使用, Yi 等人^[53]注意到特定区域的犯罪数量不仅条件依赖于该地历史犯罪事件记录, 还跟类似区域的犯罪事件记录高度相关, 使用自回归时间关联和区域间空间关联两个因素特征, 构建了一种聚类连续条件随机场(Clustered Continuous Conditional Random Field, Clustered CCRF)模型用于犯罪事件预测。为了解决 CCRF 面对区域关系稠密图的可扩展性问题, Yi 等人^[54]进一步提出了端到端的基于神经网络的 CCRF 模型, 减轻了模型训练的复杂性, 提升了整体预测模型能。

3.6 事件图谱

事件图谱(Event Graph)是通过将一系列具有逻辑关系的事件原始数据进行抽取分析与关联融合, 生成的蕴含丰富语义信息的图数据结构表示^[55]。构建好事件图谱后, 事件预测问题就可以利用一系列基于图的算法和技术来解决。本文中所指的事件图谱是广义上的概念, 只要以“图”作为事件表征工具, 就统一归为事件图谱这类技术。常见的事件图谱表示方式可以划分为以事件为节点和以事件要素为节

点两大类。

以事件为节点的事件图谱是一个有向图, 其中节点表示事件, 有向边代表事件之间的逻辑关系。将具有逻辑关系的事件按照时间序列进行串联时, 事件图谱也被成为事件序列图(Event Sequence Graph), 甚至进一步简化为事件链(Event Chain)或事件对(Event Pair), 如图 4 所示。事件序列图或事件链可以反映出事件的演化过程, 并用于事件预测分析。

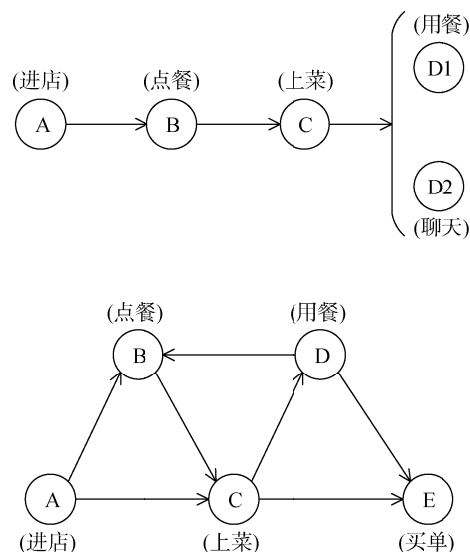


图 4 事件链与事件序列图

Figure 4 Event chain and event graph

针对事件因果关系进行事件预测的研究较为深入, 基于因果关系的事件预测系统一般包括从原始数据中抽取事件信息、构建事件间因果关系图、基于因果关系进行事件预测 3 大过程。本文旨在研究事件预测技术, 事件抽取与因果关系构建不在本文讨论范畴, 在此只讨论基于因果关系进行事件预测的研究进展, 即利用从历史事件中学习得到的事件因果关系知识, 对于给定的事件, 基于习得的因果关系来推断该事件将会导致发生的未来事件, 从而实现事件预测。Acharya 等人^[56]在实时事件流(Real-time Event Streams)场景下, 基于因果关系对接下来最有可能发生的事件进行了 top-k 预测, 且适用于有环(Cyclic)事件图谱的情况。Lei 等人^[57]在基于因果关系预测过程中, 考虑了原始数据中的上下文(Context)信息, 能够挖掘隐藏的因果关系, 进而提高事件预测的总体质量。Radinsky 等人^[58]设计实现的 Pundit 系统, 将自然语言表示的事件作为输入, 能够对该事件可能引发的未来事件进行预测, Pundit 系统的事件因果关系挖掘能力基于采集的 150 年的新闻文章和具有 200 亿条关系的知识图谱 LinkedData^[59]。

Shrestha 等人^[60]基于语义 Web 数据集构造抽象树 (Abstraction Tree) 模型, 采用聚类算法对因果事件对进行分组, 事件预测问题转化为在抽象树上遍历搜索。Zhao 等人^[61]提出了一种抽象因果关系网络 (Abstract Causality Network) 模型, 接着将其嵌入 (Embedding) 映射到连续的向量空间, 并定义一种能量函数 (Energy Function), 真正的因果事件对处于能量较低的状态, 通过最小化目标函数, 对因果事件对进行预测。Yang 等人^[62]将历史金融事件链作为输入, 同时引用金融新闻作为补充信息, 预测接下来将要发生的事件。

并不是所有的事件间的关系都可以表示为因果关系, 哈工大社会计算与信息检索研究中心 SCIR 提出了事理逻辑和事理图谱的概念^[63], 事理逻辑描述了事件之间的演化规律与模式, 事理图谱是一个事理逻辑知识库, 对现有普遍以“概念及概念间的关系”为核心的知识图谱进行了扩展, 结构上事理图谱是一个有向有环图, 节点代表事件, 有向边代表事件之间的顺承、因果、条件和上下位等逻辑关系, 具有链状、树状、环状 3 种典型子图结构。Li 等人^[64]将事理图谱应用于脚本事件预测, 即给定事件上文, 从候选事件列表中选出接下来最有可能发生的事件^[65-66]。不同于前述主要利用 Event Pair 和 Event Chain 进行预测的工作, Li 等人探索利用 Event Graph 的稠密连接信息来帮助事件预测, 构建了一个叙事事理图谱 (Narrative Event Evolutionary Graph, NEEG), 提出了可扩展的图神经网络 (Scaled Graph Neural Network, SGNN) 来解决图上的预测推理问题。

上述事件图谱或事理图谱的构建过程, 需要进行数据清洗、事件抽取和泛化、事件关系识别等大量 NLP 预处理工作, 预处理工作的误差还会向后续处理过程传导, 给后续事件预测任务带来问题。同时, 基于 Event Pair 的方法无法泛化到训练数据中不包含的因果对, 基于 Event Chain 的方法可能会丢失更早事件的信息, 带来预测模型的误差。为此, 研究者们提出了一系列“端到端”的深度学习神经网络架构模型, 基于带有记忆功能的 RNN、LSTM、GRU 及其变种等网络结构, 直接对事件预测任务的原始输入例如文本进行处理, 模型的输出为需要预测的事件语义或者描述信息^[67-71]。Magesh 等人^[72]通过实现带长短期记忆 LSTM 的 RNN 模型来实现 COVID-19 病毒传染事件预测。

深度学习模型能够通过复杂的深层网络模型从海量原始数据中自动学习丰富的事件特征知识, 然而模型内部高度的复杂性常导致人们难以理解模型

为什么输出相应的预测结果, 造成深度学习“端到端”模型的不可解释性。Deng 等人^[73]提出的动态图卷积网络模型 (Dynamic Graph Convolutional Network) 能够在预测事件的同时, 给出预测结果的上下文信息 (Context), 一定程度上解决了模型可解释性的问题。Deng 等人^[74]在此基础上, 继续对多事件及多参与主体 (Multi-event and Multi-actor) 预测问题进行了研究。Jin 等人^[75]将事件预测问题转化为时序事件图谱上的链接预测问题, 提出了一种自回归的循环事件网络 (Recurrent Event Network, RE-Net) 架构模型。Rong 等人^[76]提出的模型不仅能够揭示社交网络中的信息传播扩散过程, 而且能够解释特定社会事件发生的原因。Kapoor 等人^[77]使用图神经网络 GNN 和移动数据进行 COVID-19 预测, 在美国新冠感染数据集上进行了性能测试。

3.7 无监督方法

直观上讲, 无监督方法是从无标注的数据中学习数据的统计规律或者内在结构, 对事件进行预测的方法。Cadena 等人^[78]首先采集了 353 万多条 Twitter 数据, 使用社会学家整理的 900 多个社会内乱关键词进行过滤 (例如暴乱、抗议、示威、革命等), 并结合社交网络粉丝 (Follower)、转发等数据, 建立参与传播相关信息的级联 (Cascade) 扩散特征, 通过对此类特征进行突发性检测 (Burstiness Detection), 进而实现对社会内乱事件预测, 在巴西、墨西哥和委内瑞拉三个国家的实际数据集上进行了验证, 成功预测出了抗议等相关事件。Kattan 等人^[79]提出了一种基于遗传规划 (Genetic Programming, GP) 的无监督学习框架对指定的目标事件发生地点进行预测。Zhao 等人^[80]提出了一种动态查询扩展 (Dynamic Query Expansion) 算法, 用于迭代扩展事件领域相关关键词, 进而有效利用 Twitter 异构信息网络 (Heterogenous Information Network), 最后采用异常检测方法对可能要发生的事件进行预测。Chen 等人^[81]提出了一种基于异构社交媒体图结构的非参数异构图扫描算法 (Non-Parametric Heterogeneous Graph Scan, NPHGS), 通过最大化连通子图的非参数扫描统计量, 识别出图中的异常簇 (Cluster), 进而对异常簇的事件类型、地点、时间、参与者等信息进行预测。Shao 等人^[82]在该算法的基础上进行了扩展, 对动态多属性社交媒体网络中的子图异常检测问题^[83]进行了研究, 提出了一种新的扫描统计量函数, 基于拉格朗日松弛 (Lagrangian Relaxation) 与动态规划近似算法求解, 能够对正在进行和将要发生的事件进行预测。

3.8 多技术融合预测

顾名思义,多技术融合预测就是将上述多种事件预测技术进行有机结合,集成到一个整体事件预测系统当中,事件预测系统的典型架构如图5所示。Ramakrishnan 等人^[84]设计实现的事件自动化实时预测系统 EMBERS,融合了5种不同预测模型,能够对事件发生的时间、地点、原因以及规模等事件要素进行预测,系统还提供预测置信度的概率值。Kang 等人^[85]设计实现的 Carbon 系统与 EMBERS 类似,对抗议、罢工等社会内乱事件进行预测。Oki 等人^[86]对移动网络服务故障事件预测(Mobile Network Failure)问题进行了研究,提出了一种2层模型集成的方法,能够对基于移动网络流量、多用户 log 日志信息、Twitter 等多种数据来源集合以及逻辑回归、随机森林、自编码(AutoEncoder)等多个模型的输出进行有效集成,并在日本移动运营商的实际数据集上测试取得了较好的测试效果。Wang 等人^[87]设计了集成多种数据挖掘算法的的洪灾预测系统,能够预测未来5~15天发生的洪水灾害事件。

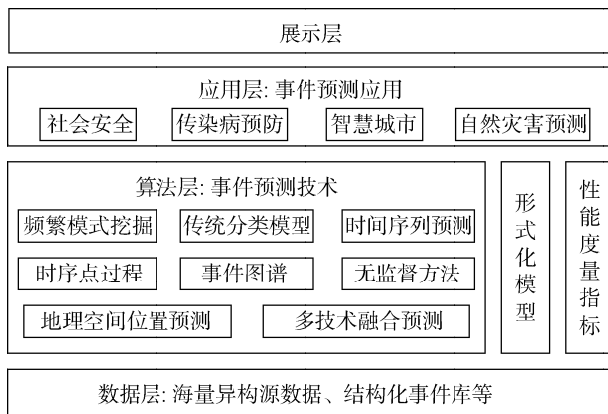


图5 事件预测系统架构

Figure 5 The architecture of event prediction system

人机结合与混合智能是目前人工智能领域研究的一大热点, Morstatter 等人^[88]设计实现的地缘政治(Geopolitical)事件预测系统 SAGE, 通过人机交互和协同, 能够更加高效地解决地缘政治事件预测这一复杂问题, 提升事件预测系统的性能。Rostami 等人^[89]将地缘政治事件预测众包(Crowdsourcing)与多任务学习相结合, 有效提升了推荐算法的性能, 进而将事件预测问题分配给对该问题预测能力更为匹配的众包参与人员, 最终提升整个系统的预测准确性。Li 等人^[90]实现的传染病预测系统也运用了群体智慧。

计划性事件预测方法也是一种采用多技术融合进行事件预测的方法, 这种方法假定绝大多数群体事件需要经过实现组织策划, 并且在组织策划的过

程中广泛运用了互联网, 在新闻、博客、社交网络平台等提前发布活动的时间、地点等相关信息。该假设在大型社会内乱事件中尤为适用, Muthiah 等人^[91]经过统计发现, 288 个群体事件中, 有 225 个(75%+)能够在新闻、社交媒体等互联网平台中找到提前组织策划线索。在这种情况下, 事件预测问题可以转化为事件检测和要素自动抽取问题, 只要能够在新闻和社交媒体等互联网内容中检测到了抗议、罢工等活动的文本、图像、视频等信息, 综合采用命名实体识别(Named Entity Recognition, NER)、光学字符识别(Optical Character Recognition, OCR)等 NLP 和 CV 技术, 进行未来时间、地点、参与者等事件要素的自动抽取, 就能够直接预测事件发生的详细信息。

3.9 事件预测技术总结和对比

每种事件预测技术都有优点和不足之处, 表1对比了本文归纳的八种事件预测技术的优点和缺点。

4 事件预测技术应用

事件预测技术作为人工智能时代的交叉学科研究领域, 其核心技术及应用模式还处于不断探索和扩展阶段。本节选取了事件预测技术对社会经济民生产生最重要影响的几个前沿应用范例, 帮助大家深刻理解事件预测技术的行业应用价值。

4.1 社会安全

社会安全事件^[5, 17-18, 22, 50-52, 73-74, 76]包括群体性事件^[6, 16, 27, 78, 80, 91-92]、恐怖主义事件^[25, 28, 93-97]、刑事犯罪案件^[98-101]等对严重影响社会安定、人民生活和经济发展的事件, 现有的应急管理系统主要应对已经发生的事件, 属于事后应急响应和处置。如果能够提前对社会安全事件进行预测和预警, 对于减少突发事件造成的生命和财产损失具有重要的意义。

恐怖主义事件是一类通过极端暴力方式制造大范围恐慌, 以达到政治、经济或者宗教目的的社会安全事件。由于恐怖主义事件具有极大的破坏性, 一直受到国内外广泛研究关注, 近年来大数据和人工智能的蓬勃发展也为此类研究提供了新技术和新思路, 已有不少文献从不同视角对恐怖主义事件预测问题进行探索^[42, 93-97], 通过总结发现历史恐怖主义事件时空规律的细微和动态特征, 对未来恐怖主义事件的空间分布和发生时机进行预测。对恐怖主义事件进行预测对探寻规律、制订有效反恐行动方案等具有重大价值。

用算法预测犯罪事件制止犯罪行为也是目前的研究热点^[41, 43, 53, 99-100], 洛杉矶警局通过与圣克拉拉大学、加州大学洛杉矶分校等合作^[101], 根据某地区

历史犯罪事件数据,通过算法预测犯罪事件最有可能发生的地点、时间段以及犯罪类型。进而根据预测结果部署相关警力,显著降低了相关区域的犯罪率。Rumi 等人^[98]对盗窃事件的风险进行了预测。

美国、欧盟近年来还资助了大量项目进行社会冲突、军事叛乱、恐怖主义活动等事件的监控和预测,构建了 EMBERS^[5, 84]、SAGE^[88]等一系列社会安

全事件预警预测系统。基于社交媒体等互联网开源信息的事件预测正在受到越来越多的关注,美国情报高级研究计划局(IARPA)资助的 OSI 项目(Open Source Indicators)成功预测巴西之春(Brazilian Spring)、委内瑞拉暴力示威等事件。IARPA 作为未来情报科技的风向标,部署的项目具有高度的战略前瞻性和参考价值,值得我们跟进深入研究。

表 1 事件预测技术总结和对比

Table 1 The summary of event prediction techniques

技术类别	优点	缺点	参考文献
频繁模式挖掘	能够同时解决多类别事件预测和事件样本类别不平衡的问题	只能预测具备特定模式的事件	[9-13]
传统分类模型	技术实现相对简单,机器学习中分类算法研究相对成熟	对训练样本的类别平衡性和质量要求较高,且不是所有事件预测问题都可建模为分类问题	[14-24]
时间序列预测	模型相对简单,需要数据样本较少,能够按照一定的趋势对事件进行预测	一般假设序列内数据点的时间间隔等长,引入了事件时间序列聚合过程中的精确时间戳信息丢失问题,外部因素变化较大时会带来较大误差	[25-30]
时序点过程	可以解决异步连续时间域的事件预测问题	强度函数的假设可能不符合实际情况,影响模型预测效果,深度点过程的可解释性较差	[31-38]
地理空间位置预测	能够对事件发生的地点进行预测	时空预测模型相对复杂,预测的准确度仍然有待提升	[39-54]
事件图谱	推理能力强,具有一定的可解释性,能够直观、清晰、准确表达事件的各属性信息	需要大量训练数据,耗时间较长	[55-77]
无监督方法	无需人工标注和整理数据,成本低	预测准确度相对较低	[78-83]
多技术融合预测	结合多种方法的优点	模型算法集成周期长,系统复杂性高	[84-91]

4.2 传染病预防

流行病(Epidemic)指短时间内可以感染人群中众多人口的疾病,传染病(Infected Disease)是由各种病原体引起的,能够在人与人、动物与动物、人与动物之间互相传播的疾病。流行病不都是传染病,但是传染病有极大可能带来全国或全球性的大流行(Pandemic),我国俗称的“瘟疫”即流行性传染病。传染病已经成为人类在地球面临的巨大威胁之一,如何预防传染病的大规模流行已经多次名列“人类社会十大科学问题”,因此对传染病爆发和传播等事件进行预测具有重要的研究价值。

新冠病毒(SARS-CoV-2)在全球范围内的大流行(COVID-19)吸引了大量研究人员的注意,涌现了传统曲线拟合、SIR 模型及 SIS/SIRD/SEIR 等变种、基于机器学习模型等众多预测方法^[72, 102-104]。因新冠疫情改为线上举行的人工智能顶级学术会议 NeurIPS 2020,收录了来自 Google Cloud AI 的论文,Arik 等人^[105]将流行病学的 SEIR 模型与机器学习模型进行

集成,预测 COVID-19 事件的发展,所提出的模型也具备良好的可解释性。Kapoor 等人^[77]使用图神经网络 GNN 和移动数据进行 COVID-19 预测,Adhikari 等人^[106]提出了一种基于 LSTM 的深度神经网络模型 EpiDeep 对流感类型的传染病爆发进行预测。斯坦福大学 Leskovec 等人^[107]在 Nature 上发表的模型预测显示少数超级传播者(Superspreader)运动轨迹中的兴趣点(Points of Interest, POI)是绝大多数传播感染所在地,对 POI 进行最大人数限制比统一均匀降低所有人员的流动性更为有效,文章还发现了不同种族和不同社会经济团体在病毒感染上的不公平性,可以为更高效和公平的预防措施政策制定提供宝贵的参考。

著名历史学家麦克尼尔曾指出,“传染病是决定人类历史进程的一个重要因素”。鼠疫、霍乱、天花、流感、艾滋病、埃博拉、新冠肺炎……人类同传染病的斗争,永无止境。从某种意义上说,人类文明史就是一部不断与瘟疫抗争的历史。事件预测技术应用于传染病预防,值得我们持续深入研究。

4.3 智慧城市

智慧城市的概念自提出以来,在国际上引起广泛关注。智慧城市已经成为提升国家治理水平、发展数字经济、提高人民生活质量的战略选择。智慧城市涵盖了智慧交通、智慧家庭等,在智慧交通领域,利用移动运营商、GPS等多种类型数据,事件预测技术可以用于预测交通事故^[44, 108]、事故持续时间^[109]、人群聚集区域^[38, 45, 47, 48, 110]等。事件预测技术可以减少交通事故和交通拥堵,为公众出行提供便捷支撑服务。在智慧家庭领域^[111],利用事件预测技术创造高效、舒适、安全、便捷的个性化家居生活。

4.4 自然灾害预测

地震、海啸、洪水、火山爆发^[35]、极端天气^[40]和森林火灾都是难以控制且破坏极大的自然灾害,事件预测技术的进展有助于预测灾害的发生。众所周知,要预报地震、火山爆发之类的灾害极其困难,但研究人员正在竭尽所能地预测余震事件,Google联合哈佛大学的研究人员在Nature发表研究成果^[112],建立了一个基于深度学习的余震事件预测模型,能够预测大地震后长达一年的余震位置。Wang等人^[87]设计了一个端到端的洪灾预测框架,能够预测未来5~15天发生的洪水灾害事件。

5 挑战与研究展望

现有的事件预测技术解决了一些应用存在的困难,然而事件预测技术的研究还处于不断发展的阶段,仍然有许多问题和挑战亟待解决,本小节列出了一些值得进一步深入研究的问题。

(1) 基于时序点过程的事件预测

目前,基于时间序列的事件预测模型已经被现有文献广泛研究,传统时间序列模型一般将事件序列转化成多段等长的时间间隔,而时序点过程将每一个事件的实际发生时间戳用于建模,已经成为解决连续时间域上的异步事件序列预测问题的有效方法。时序点过程当前的进展可以分为统计点过程和深度点过程两大方向,统计点过程建模方法具有多种强度函数形式,参数化过程中存在人工选择错误的风险,随着神经网络的发展,点过程和深度学习技术的融合展现出了强大的能力,通过大规模数据,深度点过程通过学习能够拟合更具表达能力的模型,并减少对先验知识的依赖,为事件预测打开了一扇新的大门,基于点过程的深度学习有望成为事件预测领域里的一个前沿引领的研究方向。

(2) 具备可解释性的事件预测模型

基于RNN、LSTM等深度神经网络架构的事件预测模型得到了研究人员的广泛关注^[67-71],然而现在的神经网络没有办法以一种从人类角度完全理解的方式来解释为何得到相应的事件预测结果,神经网络中的节点和边一般不具有跟事件相关的现实物理意义。Deng等人^[73, 74]提出的基于动态知识图谱的方法能够给出预测结果的上下文信息,从而辅助人工校对和决策,在模型可解释性方面进行了一些探索。我们还注意到,在人工智能顶级会议AAAI 2021上,史蒂文森理工学院和乔治梅森大学的研究人员组织了关于社会事件预测可解释性的tutorial^[113],对其基础、方法与应用等内容进行了梳理与介绍。在人工智能时代,具备可解释性的事件预测模型一定能吸引更多的研究者进行开拓,进而推动事件预测技术的精细化可解释性建模的发展。

(3) 基于多模态信息融合的事件预测

互联网上具备海量开源异构多模态信息,除了从文本中获取事件信息和相关特征外,图像、音频、视频等多种类型数据为事件分析和预测提供了更加丰富的素材。基于多模态信息融合的事件预测能够赋予计算机理解多源异构海量数据的能力,具有重要研究价值,目前已有单独利用视频数据进行事件预测的研究^[114],Lei等人^[115]给出了一个视频与字幕结合的多模态事件预测数据集VLEP,以期吸引更多基于多模态信息的事件预测研究。针对多模态数据的事件预测具有很大的研究空间,事件预测在未来发展中应融合上述多模态的信息,为预测现实世界提供更全面的视角。另外,采用有监督的学习方法往往需要投入很多的人力和物力成本,我们应该探索更多基于无监督和半监督的事件预测技术。

6 结束语

事件预测分析是社会治理与内容安全领域的重要研究内容,在未来的网络社会空间中,事件预测能力代表了感知与预知全球变化的能力。本文从事件预测的形式化模型与性能度量指标出发,对事件预测关键技术的研究和发展现状进行了全面分析和总结,探讨了事件预测技术的实际应用,展望了未来研究方向。

致谢 在此,向对本文的研究工作提供帮助的老师、同学和同事们表示感谢。作者尤其感谢哈尔滨工业大学(深圳)冯山山副教授和深圳市网联安瑞网

络科技有限公司易勇博士提供的宝贵意见与建议。

参考文献

- [1] Lazer D, Pentland A, Adamic L, et al. Computational Social Science[J]. *Science*, 2009, 323(5915): 721-723.
- [2] Tong M H, Xu B, Wang S, et al. Improving Event Detection via Open-Domain Trigger Knowledge[C]. *The 58th Annual Meeting of the Association for Computational Linguistics*, 2020: 5887-5897.
- [3] Huang L F, Ji H. Semi-Supervised New Event Type Induction and Event Detection[C]. *The 2020 Conference on Empirical Methods in Natural Language Processing*, 2020: 718-724.
- [4] Leetaru Kalev and Schrodt Philip A. Gdelt: Global data on events, location, and tone, 1979–2012[C]. *ISA annual convention*, 2013: 1-49.
- [5] Muthiah S, Butler P, Khandpur R P, et al. EMBERS At 4 Years: Experiences Operating an Open Source Indicators Forecasting System[C]. *The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016: 205-214.
- [6] Muthiah S, Huang B, Arredondo J, et al. Capturing Planned Protests from Open Source Indicators[J]. *AI Magazine*, 2016, 37(2): 63-75.
- [7] Allan J, Papka R, Lavrenko V. On-Line New Event Detection and Tracking[C]. *The 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 1998: 37-45.
- [8] Han Jiawei, Kamber Micheline and Pei Jian. Data mining concepts and techniques third edition[M]. The Morgan Kaufmann Series in Data Management Systems. 2011.
- [9] Vilalta R, Ma S. Predicting Rare Events In Temporal Domains[C]. *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, 2002: 474-481.
- [10] Laxman S, Sastry P S, Unnikrishnan K P. Discovering Frequent Episodes and Learning Hidden Markov Models: A Formal Connection[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(11): 1505-1517.
- [11] Laxman S, Tankasali V, White R W. Stream Prediction Using a Generative Model Based on Frequent Episodes In Event Sequences[C]. *The 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008: 453-461.
- [12] Zhou C, Cule B, Goethals B. A Pattern Based Predictor for Event Streams[J]. *Expert Systems With Applications*, 2015, 42(23): 9294-9306.
- [13] Letham B, Rudin C, McCormick T H, et al. Interpretable Classifiers Using Rules and Bayesian Analysis: Building a Better Stroke Prediction Model [J]. *The Annals of Applied Statistics*, 2015, 9(3): 1350-1371.
- [14] Ballings M, van den Poel D. Customer Event History for Churn Prediction: How Long is Long Enough? [J]. *Expert Systems With Applications*, 2012, 39(18): 13517-13522.
- [15] De Caigny A, Coussement K, Bock K W D. A New Hybrid Classification Algorithm for Customer Churn Prediction Based on Logistic Regression and Decision Trees[J]. *European Journal of Operational Research*, 2018, 269(2): 760-772.
- [16] Kallus N. Predicting Crowd Behavior with Big Public Data[C]. *The 23rd International Conference on World Wide Web*, 2014: 625-630.
- [17] Korkmaz G, Cadena J, Kuhlman C J, et al. Combining Heterogeneous Data Sources for Civil Unrest Forecasting[C]. *The 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 2015: 258-265.
- [18] Zhao L, Ye J P, Chen F, et al. Hierarchical Incomplete Multi-Source Feature Learning for Spatiotemporal Event Forecasting[C]. *The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016: 2085-2094.
- [19] Santiso S, Casillas A, Perez A, et al. Adverse Drug Event Prediction Combining Shallow Analysis and Machine Learning[C]. *The 5th International Workshop on Health Text Mining and Information Analysis (Louhi)*, 2014: 85-89.
- [20] Shin S Y, Kim H J. Autoencoder-Based One-Class Classification Technique for Event Prediction[C]. *The 2019 4th International Conference on Cloud Computing and Internet of Things*, 2019: 54-58.
- [21] Tops Hannah, Kunneman Florian and van den Bosch Antal. Predicting time-to-event from Twitter messages[C]. *BNAC 2013: Proceedings of the 25th Benelux Conference on Artificial Intelligence*, 2013: 8-16.
- [22] Gao Yuyang and Zhao Liang. Incomplete label multi-task ordinal regression for spatial event scale forecasting[C]. *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018: 2999-3006.
- [23] Tama B A, Comuzzi M. An Empirical Comparison of Classification Techniques for next Event Prediction Using Business Process Event Logs[J]. *Expert Systems With Applications*, 2019, 129: 233-245.
- [24] Arias M, Arratia A, Xuriguera R. Forecasting with Twitter Data[J]. *ACM Transactions on Intelligent Systems and Technology*, 2014, 5(1): 1-24.
- [25] Yonamine James E. Predicting future levels of violence in afghanistan districts using gdelt. <http://data.gdeltproject.org/documentation/Predicting-Future-Levels-of-Violence-in-Afghanistan-Districts-using-GDELTPdf>. 2013.
- [26] Povinelli R J, Feng X. A New Temporal Pattern Identification Method for Characterization and Prediction of Complex Time Series Events[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2003, 15(2): 339-352.
- [27] Zhao L, Chen F, Lu C T, et al. Spatiotemporal Event Forecasting In Social Media[C]. *The 2015 SIAM International Conference on Data Mining*, 2015: 963-971.
- [28] Petroff V B, Bond J H, Bond D H, et al. Using Hidden Markov Models to Predict Terror before it Hits (again) [J]. *New York, NY*, 2013:163-180.
- [29] Salfner F, Malek M. Using Hidden Semi-Markov Models for Effective Online Failure Prediction[C]. *2007 26th IEEE International Symposium on Reliable Distributed Systems*, 2007: 161-174.
- [30] Antunes M, Turkman M A A, Turkman K F. A Bayesian Approach to Event Prediction[J]. *Journal of Time Series Analysis*, 2003, 24(6): 631-646.
- [31] Minor B, Cook D J. Forecasting Occurrences of Activities[J]. *Pervasive and Mobile Computing*, 2017, 38: 77-91.
- [32] Ertekin Ş, Rudin C, McCormick T H. Reactive Point Processes: A

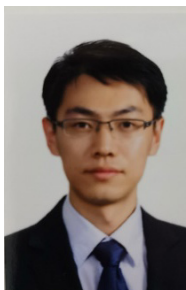
- New Approach to Predicting Power Failures In Underground Electrical Systems[J]. *The Annals of Applied Statistics*, 2015, 9(1): 122-144.
- [33] Simma A, Jordan M I. Modeling Events with Cascades of Poisson Processes [EB/OL]. 2012: arXiv preprint arXiv:1203.3516.
- [34] Du N, Dai H J, Trivedi R, et al. Recurrent Marked Temporal Point Processes: Embedding Event History to Vector[C]. *The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016: 1555-1564.
- [35] Damaschke M, Cronin S J, Bebbington M S. A Volcanic Event Forecasting Model for Multiple Tephra Records, Demonstrated on Mt. Taranaki, New Zealand[J]. *Bulletin of Volcanology*, 2017, 80(1): 1-14.
- [36] Qiao Z, Zhao S W, Xiao C, et al. Pairwise-Ranking Based Collaborative Recurrent Neural Networks for Clinical Event Prediction[C]. *The Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2018: 3520-3526.
- [37] Dempsey W H, Moreno A, Scott C K, et al. ISurvive: An Interpretable, Event-Time Prediction Model for mHealth [J]. *The Machine Learning Research*, 2017, 70(970).
- [38] Vahedian A, Zhou X, Tong L, et al. Predicting Urban Dispersal Events: A Two-Stage Framework through Deep Survival Analysis on Mobility Data[J]. *The AAAI Conference on Artificial Intelligence*, 2019, 33: 5199-5206.
- [39] Jiang Z. A Survey on Spatial Prediction Methods[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 31(9): 1645-1664.
- [40] Wang D W, Ding W. A Hierarchical Pattern Learning Framework for Forecasting Extreme Weather Events[C]. *2015 IEEE International Conference on Data Mining*, 2015: 1021-1026.
- [41] Xiong C X, Srivastava A, Kannan R, et al. On Predicting Crime with Heterogeneous Spatial Patterns: Methods and Evaluation[C]. *The 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2019: 43-51.
- [42] Hao M M, Jiang D, Ding F Y, et al. Simulating Spatio-Temporal Patterns of Terrorism Incidents on the Indochina Peninsula with GIS and the Random Forest Method[J]. *ISPRS International Journal of Geo-Information*, 2019, 8(3): 133.
- [43] Ristea A, Al Boni M, Resch B, et al. Spatial Crime Distribution and Prediction for Sporting Events Using Social Media[J]. *International Journal of Geographical Information Science*, 2020, 34(9): 1708-1739.
- [44] Bao J, Liu P, Ukkusuri S V. A Spatiotemporal Deep Learning Approach for Citywide Short-Term Crash Risk Prediction with Multi-Source Data[J]. *Accident Analysis & Prevention*, 2019, 122: 239-254.
- [45] Mukhina K D, Vishneratin A A, Nasonov D. Urban Events Prediction via Convolutional Neural Networks and Instagram Data[J]. *Procedia Computer Science*, 2019, 156: 176-184.
- [46] Wang B, Yin P H, Bertozzi A L, et al. Deep Learning for Real-Time Crime Forecasting and Its Ternarization[J]. *Chinese Annals of Mathematics, Series B*, 2019, 40(6): 949-966.
- [47] Vahedian A, Zhou X, Tong L, et al. Forecasting Gathering Events through Continuous Destination Prediction on Big Trajectory Data[C]. *The 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2017: 1-10.
- [48] Khezerlou A V, Zhou X, Tong L, et al. Forecasting Gathering Events through Trajectory Destination Prediction: A Dynamic Hybrid Model[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(3): 991-1004.
- [49] Zhao L, Sun Q, Ye J P, et al. Multi-Task Learning for Spatio-Temporal Event Forecasting[C]. *The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015: 1503-1512.
- [50] Zhao L, Sun Q, Ye J P, et al. Feature Constrained Multi-Task Learning Models for Spatiotemporal Event Forecasting[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(5): 1059-1072.
- [51] Zhao L, Wang J X, Chen F, et al. Spatial Event Forecasting In Social Media with Geographically Hierarchical Regularization[J]. *Proceedings of the IEEE*, 2017, 105(10): 1953-1970.
- [52] Ning Y, Tao R R, Reddy C K, et al. STAPLE: Spatio-Temporal Precursor Learning for Event Forecasting[M]. *Proceedings of the 2018 SIAM International Conference on Data Mining*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2018: 99-107.
- [53] Yi F, Yu Z W, Zhuang F Z, et al. An Integrated Model for Crime Prediction Using Temporal and Spatial Factors[C]. *2018 IEEE International Conference on Data Mining*, 2018: 1386-1391.
- [54] Yi F, Yu Z W, Zhuang F Z, et al. Neural Network Based Continuous Conditional Random Field for Fine-Grained Crime Prediction[C]. *The Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2019: 4157-4163.
- [55] Glavaš G, Šnajder J. Construction and Evaluation of Event Graphs[J]. *Natural Language Engineering*, 2015, 21(4): 607-652.
- [56] Acharya S, Lee B S, Hines P. Causal Prediction of Top-k Event Types over Real-Time Event Streams[J]. *The Computer Journal*, 2017, 60(11): 1561-1581.
- [57] Lei L, Ren X G, Franciscus N, et al. Event Prediction Based on Causality Reasoning[M]. *Intelligent Information and Database Systems*. Cham: Springer International Publishing, 2019: 165-176.
- [58] Radinsky K, Davidovich S, Markovitch S. Learning Causality for News Events Prediction[C]. *The 21st international conference on World Wide Web*, 2012: 909-918.
- [59] Bizer C, Heath T, Berners-Lee T. Linked Data - the Story so far[J]. *International Journal on Semantic Web and Information Systems*, 2009, 5(3): 1-22.
- [60] Shrestha Prajwal, Lee Byung Suk and Bagrow James P. Predicting an Effect Event from a New Cause Event Using a Semantic Web Based Abstraction Tree of Past Cause-Effect Event Pairs[C]. *SIM-Big*, 2017: 198-207.
- [61] Zhao S D, Wang Q, Massung S, et al. Constructing and Embedding Abstract Event Causality Networks from Text Snippets[C]. *The Tenth ACM International Conference on Web Search and Data Mining*, 2017: 335-344.
- [62] Yang Y Y, Wei Z Y, Chen Q, et al. Using External Knowledge for Financial Event Prediction Based on Graph Neural Networks[C]. *The 28th ACM International Conference on Information and*

- Knowledge Management*, 2019: 2161-2164.
- [63] Ding X, Li Z Y, Liu T, et al. ELG: An Event Logic Graph [EB/OL]. 2019: arXiv preprint arXiv:1907.08015.
- [64] Li Z Y, Ding X, Liu T. Constructing Narrative Event Evolutionary Graph for Script Event Prediction [EB/OL]. 2018: arXiv preprint arXiv:1805.05081.
- [65] Granroth-Wilding Mark and Clark Stephen. What happens next? event prediction using a compositional neural network model[C]. *The Thirtieth AAAI Conference on Artificial Intelligence*, 2016: 2727-2733.
- [66] Lv S W, Qian W H, Huang L T, et al. SAM-Net: Integrating Event-Level and Chain-Level Attentions to Predict what Happens next[J]. *The AAAI Conference on Artificial Intelligence*, 2019, 33: 6802-6809.
- [67] Hu L M. Integrating Hierarchical Attentions for Future Subevent Prediction[J]. *IEEE Access*, 2019, 8: 3106-3114.
- [68] Hu Linmei, Li Juanzi, Nie Liqiang, et al. What happens next? future subevent prediction using contextual hierarchical LSTM[C]. *The Thirty-First AAAI Conference on Artificial Intelligence*, 2017: 3450-3456.
- [69] Su Z C, Jiang J L. Hierarchical Gated Recurrent Unit with Semantic Attention for Event Prediction[J]. *Future Internet*, 2020, 12(2): 39.
- [70] Xue C, Zeng Z H, He Y Y, et al. A MIML-LSTM Neural Network for Integrated Fine-Grained Event Forecasting[C]. *The 2018 International Conference on Big Data Technologies - ICBTD '18*, 2018: 44-51.
- [71] Yu S Q, Hu L M, Wu B. DRAM: A Deep Reinforced Intra-Attentive Model for Event Prediction[M]. *Knowledge Science, Engineering and Management*. Cham: Springer International Publishing, 2019: 701-713.
- [72] Magesh S, Niveditha V R, Rajakumar P S, et al. Pervasive Computing In the Context of COVID-19 Prediction with AI-Based Algorithms[J]. *International Journal of Pervasive Computing and Communications*, 2020, 16(5): 477-487.
- [73] Deng S, Rangwala H, Ning Y. Learning Dynamic Context Graphs for Predicting Social Events[C]. *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 1007-1016.
- [74] Deng Songgaojun, Rangwala Huzefa and Ning Yue. Dynamic Knowledge Graph based Multi-Event Forecasting[C] *The 26th KDD 2020: Virtual Conference*, 2020:1585-1595.
- [75] Jin W, Qu M, Jin X S, et al. Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs[C]. *The 2020 Conference on Empirical Methods in Natural Language Processing*, 2020: 6669-6683.
- [76] Rong Y, Cheng H, Mo Z Y. Why it Happened: Identifying and Modeling the Reasons of the Happening of Social Events[C]. *The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015: 1015-1024.
- [77] Kapoor A, Ben X, Liu L Y, et al. Examining COVID-19 Forecasting Using Spatio-Temporal Graph Neural Networks[EB/OL]. 2020: arXiv preprint arXiv:2007.03113.
- [78] Cadena J, Korkmaz G, Kuhlman C J, et al. Forecasting Social Unrest Using Activity Cascades[J]. *PLoS One*, 2015, 10(6): e0128879.
- [79] Kattan A, Fatima S, Arif M. Time-Series Event-Based Prediction: An Unsupervised Learning Framework Based on Genetic Programming[J]. *Information Sciences*, 2015, 301: 99-123.
- [80] Zhao L, Chen F, Dai J, et al. Unsupervised Spatial Event Detection In Targeted Domains with Applications to Civil Unrest Modeling[J]. *PLoS One*, 2014, 9(10): e110206.
- [81] Chen F, Neill D B. Non-Parametric Scan Statistics for Event Detection and Forecasting In Heterogeneous Social Media Graphs[C]. *The 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014: 1166-1175.
- [82] Shao M L, Li J X, Chen F, et al. An Efficient Approach to Event Detection and Forecasting In Dynamic Multivariate Social Media Networks[C]. *The 26th International Conference on World Wide Web*, 2017: 1631-1639.
- [83] Chen F, Zhou B J, Alim A, et al. A Generic Framework for Interesting Subspace Cluster Detection In Multi-Attributed Networks[C]. *2017 IEEE International Conference on Data Mining*, 2017: 41-50.
- [84] Ramakrishnan N, Butler P, Muthiah S, et al. 'Beating the News' with EMBERS: Forecasting Civil Unrest Using Open Source Indicators[C]. *The 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014: 1799-1808.
- [85] Kang W, Chen J, Li J Y, et al. Carbon: Forecasting Civil Unrest Events by Monitoring News and Social Media [C]. 2017: 859-865.
- [86] Oki Motoyuki, Takeuchi Koh and Uematsu Yukio. Mobile network failure event detection and forecasting with multiple user activity data sets[C]. *AAAI*, 2018: 7786-7792.
- [87] Wang D W, Ding W, Yu K, et al. Towards Long-Lead Forecasting of Extreme Flood Events: A Data Mining Framework for Precipitation Cluster Precursors Identification[C]. *The 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013: 1285-1293.
- [88] Morstatter F, Galstyan A, Satyukov G, et al. SAGE: A Hybrid Geopolitical Event Forecasting System[C]. *The Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2019: 6557-6559.
- [89] Rostami M, Huber D, Lu T C. A Crowdsourcing Triage Algorithm for Geopolitical Event Forecasting[C]. *The 12th ACM Conference on Recommender Systems*, 2018: 377-381.
- [90] Li E Y, Tung C Y, Chang S H. The Wisdom of Crowds In Action: Forecasting Epidemic Diseases with a Web-Based Prediction Market System[J]. *International Journal of Medical Informatics*, 2016, 92: 35-43.
- [91] Muthiah Sathappan, Huang Bert, Arredondo Jaime, et al. Planned protest modeling in news and social media[C]. *Twenty-Seventh IAAI Conference*, 2015: 3920-3927.
- [92] Qiao F C, Li P, Zhang X, et al. Predicting Social Unrest Events with Hidden Markov Models Using GDELT[J]. *Discrete Dynamics in Nature and Society*, 2017, 2017: 1-13.
- [93] Feng Y, Wang D J, Yin Y Q, et al. An XGBoost-Based Casualty Prediction Method for Terrorist Attacks[J]. *Complex & Intelligent Systems*, 2020, 6(3): 721-740.

- [94] Spiliotopoulos D, Vassilakis C, Margaris D. Data-Driven Country Safety Monitoring Terrorist Attack Prediction[C]. *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2019: 1128-1135.
- [95] Meng X, Nie L Y, Song J P. Big Data-Based Prediction of Terrorist Attacks[J]. *Computers & Electrical Engineering*, 2019, 77: 120-127.
- [96] ALfatih M, Li C L, Saadalla N E. Prediction of Groups Responsible for Terrorism Attack Using Tree Based Models[C]. *The 2019 International Conference on Artificial Intelligence and Computer Science*, 2019: 320-324.
- [97] Agarwal P, Sharma M, Chandra S. Comparison of Machine Learning Approaches In the Prediction of Terrorist Attacks[C]. *2019 Twelfth International Conference on Contemporary Computing*, 2019: 1-7.
- [98] Rumi S K, Deng K, Salim F D. Theft Prediction with Individual Risk Factor of Visitors[C]. *The 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2018: 552-555.
- [99] Ensign Danielle, Friedler Sorelle A., Neville Scott, et al. Runaway Feedback Loops in Predictive Policing[C]. *The 1st Conference on Fairness, Accountability and Transparency*, 2018: 160-171.
- [100] Shapiro A. Reform Predictive Policing[J]. *Nature*, 2017, 541(7638): 458-460.
- [101] Mohler G O, Short M B, Malinowski S, et al. Randomized Controlled Field Trials of Predictive Policing[J]. *Journal of the American Statistical Association*, 2015, 110(512): 1399-1411.
- [102] Achterberg M A, Prasse B, Ma L, et al. Comparing the Accuracy of Several Network-Based COVID-19 Prediction Algorithms[J]. *International Journal of Forecasting*, 2020: Article in Press.
- [103] Anastassopoulou C, Russo L, Tsakris A, et al. Data-Based Analysis, Modelling and Forecasting of the COVID-19 Outbreak[J]. *PLoS One*, 2020, 15(3): e0230405.
- [104] Petropoulos F, Makridakis S. Forecasting the Novel Coronavirus COVID-19[J]. *PLoS One*, 2020, 15(3): e0231236.
- [105] Arik S O, Li C L, Yoon J, et al. Interpretable Sequence Learning for COVID-19 Forecasting [J]. *Advances in Neural Information Processing Systems*, 2020, 33.
- [106] Adhikari B, Xu X F, Ramakrishnan N, et al. EpiDeep: Exploiting Embeddings for Epidemic Forecasting[C]. *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 577-586.
- [107] Chang S, Pierson E, Koh P W, et al. Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening[J]. *Nature*, 2021, 589(7840): 82-87.
- [108] Yuan Z N, Zhou X, Yang T B. Hetero-ConvLSTM: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-Temporal Data[C]. *The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018: 984-992.
- [109] Fu K Q, Ji T R, Zhao L, et al. TITAN: A Spatiotemporal Feature Learning Framework for Traffic Incident Duration Prediction[C]. *The 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2019: 329-338.
- [110] Jiang R H, Song X, Huang D, et al. DeepUrbanEvent: A System for Predicting Citywide Crowd Dynamics At Big Events[C]. *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 2114-2122.
- [111] Casagrande F D, Tørresen J, Zouganeli E. Sensor Event Prediction Using Recurrent Neural Network In Smart Homes for Older Adults[C]. *2018 International Conference on Intelligent Systems*, 2018: 662-668.
- [112] DeVries P M R, Viégas F, Wattenberg M, et al. Deep Learning of Aftershock Patterns Following Large Earthquakes[J]. *Nature*, 2018, 560(7720): 632-634.
- [113] Deng Songgaojun, Ning Yue and Rangwala Huzefa. Explainable AI for Societal Event Predictions: Foundations, Methods, and Applications. <https://yue-ning.github.io/aaai-21-tutorial.html>. 2021.
- [114] Neumann L, Zisserman A, Vedaldi A. Future Event Prediction: If and when[C]. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019: 2935-2943.
- [115] Lei J, Yu L C, Berg T, et al. What is more Likely to Happen Next? Video-and-Language Future Event Prediction[C]. *The 2020 Conference on Empirical Methods in Natural Language Processing*, 2020: 8769-8784.



李志鹏 于 2013 年在中国科学技术大学通信与信息系统专业获得硕士学位。现在中国科学技术大学网络空间安全专业攻读博士学位, 现任深圳市网联安瑞网络科技有限公司工程师。研究领域为数据挖掘、网络安全。研究兴趣包括: 事件预测、人工智能系统安全。Email: lzp@mail.ustc.edu.cn



杨阳朝 于 2014 年在中国科学技术大学计算机应用专业获得博士学位。现任深圳市网联安瑞网络科技有限公司高级工程师。研究领域为网络空间内容与认知安全。研究兴趣包括: 人工智能、认知对抗、社交网络等。Email: forester@mail.ustc.edu.cn



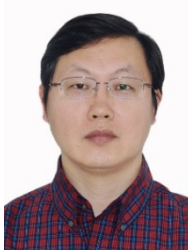
廖勇 于 2009 年在美国麻省大学 Amherst 分校计算机工程专业获得博士学位。现为中国科学技术大学网络空间安全学院客座教授, 现任深圳市网联安瑞网络科技有限公司研究员。研究领域为数据科学、网络安全。研究兴趣包括: 网络测量、网络系统安全、数据分析等。Email: ly@ustc.edu



俞能海 于 2004 年在中国科学技术大学信息与通信工程专业获得博士学位。现任中国科学技术大学网络空间安全学院执行院长。研究领域为视频处理与多媒体通信、信息检索与数据挖掘、媒体内容安全等。Email: ynh@ustc.edu.cn



吴哲 于 2019 年在西安建筑科技大学城市规划专业获得硕士学位。现任深圳市网联安瑞网络科技有限公司助理工程师。研究领域为数据挖掘、智慧城市。研究兴趣包括: 机器学习、事件预测、数据可视化。Email: 956693827@qq.com



谢海永 于 2007 年在美国耶鲁大学计算机专业获得博士学位。现任中国科学技术大学网络空间安全学院教授、博士生导师。研究领域为网络空间科学, 研究兴趣包括网络体系结构、网络与系统等。Email: hxie@ustc.edu.cn



石璐 于 2017 年在中国科学院大学计算机软件与理论专业获得博士学位。现任深圳市网联安瑞网络科技有限公司高级工程师。研究领域为网络安全、网络空间内容与认知域安全, 研究兴趣包括: 认知图谱、社交网络分析。Email: jshi@nscslab.net



曾曦 于 2008 年在电子科技大学软件工程专业获得硕士学位。现任深圳市网联安瑞网络科技有限公司研究员。研究领域为网络空间认知域安全、网络通信。研究兴趣为综合交叉学科应用、工程体系设计。Email: zxmm2@163.com