犯罪数据可视化研究进展

丁伟杰[1,2,3], 梁荣华[4], 孙国道[4)*], 柴叶[5], 王勋[5], 郑滋椀[2,3]

[1] (浙江工业大学信息工程学院 杭州 310023)
[2] (浙江警察学院大数据与网络安全研究院 杭州 310053)
[3] (基于大数据架构的公安信息化应用公安部重点实验室 杭州 310053)
[4] (浙江工业大学计算机科学与技术学院 杭州 310023)
[5] (浙江警察学院计算机与信息安全系 杭州 310053)
(guodao@zjut.edu.cn)

摘 要: 犯罪数据分析与挖掘已成为大数据时代犯罪学研究及现代社会治理的基础手段, 犯罪数据挖掘在揭示犯罪模式、分析犯罪原因、评估犯罪防控效果等方面具有重要作用. 首先, 以犯罪模式发现与预测、警力资源配置与优化、案件侦查分析与研判 3 种业务为导向, 对犯罪数据可视分析任务进行归纳, 系统地梳理了犯罪数据可视化领域的研究与应用现状; 其次, 针对犯罪时空数据、犯罪文本数据、犯罪网络关系数据等不同类型数据的可视化方式和技术手段, 展开了系统性的分析; 最后, 总结和展望了犯罪数据可视化面临的挑战和发展方向.

关键词: 犯罪数据; 可视化; 犯罪模式; 研究进展
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023.19566

Research Progress of Crime Data Visualization

Ding Weijie[1,2,3], Liang Ronghua[4], Sun Guodao[4)*], Chai Ye[5], Wang Xun[5], and Zheng Ziwan[2,3]

[1] (*College of Information Engineering, Zhejiang University of Technology, Hangzhou* 310023)
[2] (*Big-Data and Network Security Research Institute, Zhejiang Police College, Hangzhou* 310053)
[3] (*Key Laboratory of Public Security Information Application Based on Big-Data Architecture, Ministry of Public Security, Hangzhou* 310053)
[4] (*College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou* 310023)
[5] (*Department of Computer and Information Security, Zhejiang Police College, Hangzhou* 310053)

Abstract: Crime data analysis and mining have become the basic means of criminology research and modern social governance in the era of big data. Crime data plays an irreplaceable role in revealing the crime pattern, analyzing the crime causes and evaluating the effect of crime prevention. Firstly, guided by three business orientations including crime pattern discovery and prediction, police resource allocation and optimization, case investigation analysis and judgment, we summarize the crime data visual analysis tasks, systematically combed the research and application status of crime data visualization. Secondly, the paper also conducts systematic literature research on the visual analysis methods and technical means of different types of data, such as crime spatio-temporal data, crime text data and crime network data. Finally, the challenges of crime data visualization are summarized, as well as prospecting the development.

收稿日期: 2022-01-06 修回日期: 2022-05-07. **基金项目**: 国家自然科学基金重点项目(62036009); 教育部人文社会科学研究规划课题(22YJA840004); 公安部科技强警基础工作专项(2020GABJC35); 国家级创新创业训练计划(202111483005). **丁伟杰**(1981—), 男, 博士研究生, 副教授, 主要研究方向为犯罪数据可视分析; **梁荣华**(1974—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究方向为可视化和计算机图形图像处理; **孙国道**(1988—), 男, 博士, 副教授, 论文通信作者, 主要研究方向为信息可视化与可视分析; **柴叶**(2001—), 女, 2019 级本科生, 研究方向为警务数据分析; **王勋**(1988—), 男, 讲师, 主要研究方向为数据可视化; **郑滋椀**(1989—), 男, 实验师, 主要研究方向为警务数据分析.

Key words: crime data; visualization; criminal pattern; research progress

犯罪数据可视化是指数据可视化技术在犯罪数据分析与挖掘领域的具体应用. 近年来, 随着信息技术的发展, 警务信息化基础设施不断完善, 积累了海量的警务数据. 这些数据中隐含着大量的犯罪模式, 因此, 警务数据挖掘并从中发现犯罪规律至关重要. 同时, 交互式大数据可视分析技术的应用, 使得传统的描述性犯罪分析逐步向探索式犯罪分析方向转变, 尤其是预测警务在美国的兴起和成功实践, 进一步促进了犯罪分析研究的发展.

犯罪数据可视分析技术能够直观、快速地展示犯罪数据的特征, 揭示犯罪数据中蕴含的规律, 如犯罪事件的高发时间和地点等. 同时, 人机交互所提供的丰富、流畅的交互方法, 能够更加友好、全面地挖掘犯罪数据中隐含的犯罪模式, 为警务部门快速分析案件提供依据. 近年来, 犯罪数据的可视分析受到业内相关研究人员的重点关注, 产出了大批研究成果.

通过文献检索分析发现, 国外的犯罪数据可视化研究主要集中在犯罪时空分析方面, 而国内的相关研究由于警务数据的开放度不够, 导致以业务应用为导向的相关工作较少. 如仅以警务工作中常见的业务方向为犯罪数据可视化任务进行分类总结, 则难以支撑综合性的评述和论断. 因此, 本文采用“业务归纳+技术总结+业务展望”的总体架构开展综述. 首先, 以警务工作业务场景为导向, 分析归纳了犯罪数据可视化目标、犯罪数据类型、可视化方式和可视化技术分类之间的关系; 其次, 根据犯罪数据的不同属性, 分别从时空、文本和网络关系可视化 3 个角度对犯罪数据可视化技术进行归纳和总结; 最后, 从警务应用角度展望对犯罪数据可视化未来的发展与挑战. 以期能够为犯罪数据可视化研究者提供该领域研究现状的概括性总结, 拓展不具备警务业务经验的广大可视化研究人员的研究视野; 同时, 促进警务专家对可视化技术的了解, 扩大可视化技术应用在警务实务界的影响力.

1 犯罪数据类型及其可视化分析任务

1.1 犯罪数据类型及其特征

犯罪数据是指警务部门对描述与犯罪相关的人、地、事、物、组织等要素, 所采集的特征、属性、状态、趋势、关系等内容的数据, 其类型涵盖数值、文字、语音、图形、图像和视频等. 在数据可视化分析过程中, 根据数据的用途、结构等因素, 犯罪数据可以分为犯罪时空数据、犯罪文本数据和犯罪网络关系数据 3 种类型.

(1) 犯罪时空数据. 包含时空信息的各类犯罪数据分为静态时空数据和动态时空数据. 其中, 静态时空犯罪数据描述了相关要素某一时刻的空间属性, 如警情、犯罪人员住所等; 动态时空数据则描述相关要素的连续变化的时空属性, 如犯罪人员时空轨迹、车辆时空轨迹、犯罪时间序列等. 受社会、人文、地理、情感等多因素的综合影响, 犯罪时空数据具有多源异构、多尺度、时空演变性等特点, 呈现复杂、多变的状态. 例如, 与其他时序数据相比, 犯罪时间序列具有更明显的随机性和波动性.

(2) 犯罪文本数据. 警务部门采集的各类文本数据, 如简要案情、笔录、法律文书等. 这类数据包含了大量的警务要素的各类属性, 如案件的过程信息、人员的属性特征、案发地点、时间、作案方式、工具等. 涉案文本虽然具有相对固定的格式, 但具有较强的业务复杂性, 其语义表述方式因人而异, 且文本实体要素维度较多. 与其他领域文本数据相比, 该类数据具有半结构化、高维度、多语义和冗余性强等特点.

(3) 犯罪网络关系数据. 包含要素网络关联信息的犯罪数据, 如通信数据和交通行为主体之间一定规则约束下的时空伴随数据、社交网络数据、快递数据等. 根据数据能否直接、准确地反映两者之间是否存在关联, 可分为强关联数据(如通信、同行、同住等数据)和弱关联数据(如同航班、同火车数据). 其中, 强关联数据表明两者之间必定存在某种关联, 弱关联犯罪数据表明两者之间只是可能存在关联. 强关联犯罪数据和弱关联犯罪数据共同构成了犯罪人员、事件之间的网络关系. 犯罪网络的层次、结点与案件的复杂度紧密相关, 该类数据具有大体量、多中心、动态关联和稀疏度不一致等特点.

由于犯罪数据有其自身的特点, 能够全面地反映犯罪相关人、事、物的数量特征、时空分布特征以及变化过程, 犯罪数据可视化在较大程度上可以辅助警务部门开展犯罪模式发现、犯罪成因分析以及犯罪趋势挖掘等工作.

1.2 犯罪数据可视化分析任务

犯罪数据可视化是揭示犯罪规律的重要方式

和手段. 犯罪数据应用遵循数据-信息-知识-指导实践的数据应用路径, 而犯罪数据可视化正处在犯罪数据-犯罪信息-发现犯罪知识的关键环节. 犯罪知识的发现势必要与具体业务场景相结合. 例如, 犯罪时空可视化可帮助警方科学调配警力资源、优化警务巡逻方案, 科学设计巡逻路线, 展示人员时空行为轨迹等; 犯罪时序可视化可以挖掘展示犯罪种类及数量在时间维度上的变化趋势, 及时发现犯罪演变规律, 指导警务部门调整勤务机制; 犯罪网络关系可视化可以分析犯罪网络关系及其团伙组织架构、犯罪影响因子; 犯罪文本语义可视化可以帮助警方挖掘犯罪线索、串并特征相似的案件等. 这些工作的可视化目标根据工作场景归属可归纳为 3 个方面: (1) 犯罪模式发现与预测. 该工作场景较为依赖对历史犯罪数据的分析, 从犯罪案事件时间演变、时空分布、高频关键词统计、探索犯罪团伙的网络结构等角度挖掘案件特征, 可视化技术的应用涵盖了犯罪时空、网络和文本可视化; (2) 警力资源配置与优化. 该工作场景较为依赖犯罪案事件的空间分布特征, 由于警情数据同时存在密集区域和稀疏区域, 在空间上分布极不均匀. 因此, 利用可视化技术实现空间维度可用和高效的交互分析是警务地理信息系统(police geographic information system, PGIS)的重要目标; (3) 案件侦查分析与研判. 该工作场景较为重视犯罪案事件数据的过程性分析, 涵盖了犯罪文本、网络可视化手段. 可视化技术的应用可以解决案情文本冗余问题, 使案件关系人特征提取、线索挖掘、案件有效关联和串并等任务更具有可表示性. 表 1 所示为对犯罪数据可视化任务代表性工作的总结.

表 1　犯罪数据可视化任务总结

可视化目标	数据类型	可视化方式	可视化类别	文献
犯罪模式发现与预测	案件时序数据	折线图	时序可视化	[1]
	案件时空数据	交互式动画、时空立方体、地图矩阵	时空协同 Detecting criminal 可视化	[2-4]
	案件时序数据	混淆矩阵热力图	时序可视化	[5]
	案件空间数据	分级统计图	犯罪制图、空间可视化	[6-8]
	案件时空数据	平行坐标图、时间矩阵图	时空协同可视化	[9]
	案件统计数据	维诺图	空间/网络可视化	[10]
	案件统计数据、社会经济数据	平行坐标图、自组织地图、时间矩阵图	多维数据可视化	[11]
	手机通讯数据	散点图、堆栈图、数据花瓶	时序可视化	[12]
	案卷报告文本	知识图谱	文本可视化	[13]
	交通事故统计	网格距离热点	空间可视化	[14]
警力资源配置与优化	接报警数据	散点图	空间可视化	[15-17]
	PGIS 数据	道路街区热力图	空间可视化	[6]
	PGIS 数据	核密度估计	空间可视化	[18-19]
	PGIS 数据	最小凸多边形	空间可视化	[20-21]
案件侦查分析与研判	社交媒体数据、案件报告数据	立体散点图、三维柱状图	时空可视化、多维数据可视化	[22]
	案件报告数据、社交媒体数据	甘特图、多图协同	时空协同可视化	[23]
	案卷报告文本	关键词词云图	文本可视化	[24]
	案卷报告文本、接报警数据	地理时间标签	文本可视化、空间可视化	[25]
	案件报告文本	Netmap	关系图谱	[26]
	案件报告文本	SNA	网络可视化	[27-29]
	网络攻击数据	3D 柱面图、3D 散点图, 热力图	网络可视化	[30-32]
	PGIS 数据	平面平行坐标图、立体平行坐标图	多维数据可视化、空间可视化	[11,33]

2　犯罪数据可视分析方法

本文调研和探讨了近 20 年来国内外犯罪数据可视化研究成果. 借助 Web of Science、谷歌学术、中国知网和万方等中英文数据库, 以“犯罪”“警务”“安全”“数据”“分析”“可视化”等中英文关键词组合进行检索, 获取相关期刊、会议论文 100 余篇, 经过筛选得到有代表性的中英文文献 61 篇.

文献调研结果表明，犯罪时空数据可视化一直是学者研究的重点，整体研究热度稳步进行、时有突破，散点视图、分级视图和时序视图等是犯罪时空数据可视描述的主要呈现方式．2012—2014年，随着机器学习技术在犯罪文本分析中的应用，犯罪文本可视化分析研究也得到了长足发展．犯罪网络关系数据可视化研究起步较早，随着关系视图技术实现上的突破，其在近 10 年来受关注度较高．标签视图、组合视图和轨迹视图是主要的可视化呈现方式．总体来讲，以可视化实现方式为维度统计，时序视图最早受到关注，分级视图研究文献最多，而组合视图和关系视图应用范围最广，如图 1 所示，图例圆饼大小为文献数量的映射．本节将从犯罪时空数据可视化、犯罪文本数据可视化和犯罪网络关系数据可视化 3 个角度对近 20 年来本领域代表性研究成果进行总结．

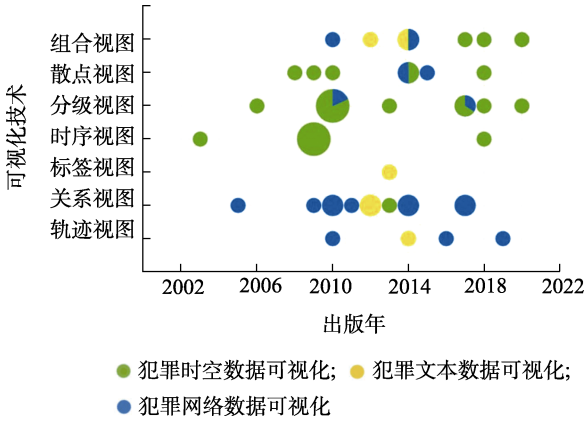


图 1 犯罪数据可视化文献统计

2.1 犯罪时空数据可视化

大量的研究表明[1,18,34-36]，犯罪行为与时间、地理环境有密切的关系，不同类型的犯罪事件具有不同的时间和空间维度特征．针对犯罪数据的时空特征，无论是数据分析人员还是执法机关，通过不同的作图方式和时空聚类分析来对犯罪时空模型进行可视化．

犯罪时空数据的可视化方法可以从基于时序的可视分析、基于空间的可视分析和基于时空协同的可视化 3 个方面展开．

2.1.1 基于时序的犯罪数据可视化

基于时序的可视分析方法通常根据统计学原理，对特定时间段犯罪事件的发生时间信息进行量化统计，从而发现犯罪事件在时间上的分布特征和趋势．传统的犯罪预测是根据统计数据模型预测特定时间和地点的犯罪数量和类型．例如，Gorr 等[1]根据 Pittsburgh 历年来不同犯罪类型的数据，采用多幅简单的折线图展示不同类型的犯罪数量在城市不同区域随时间的变化情况．基于时间的犯罪数据可视分析除了展示数据量的规模，还能够提供更多的有效信息．例如，Ferrara 等[37]对犯罪嫌疑人的手机通讯次数、通讯类型在时间维度进行统计分析，采用图 2a 所示事件流散点图来揭示犯罪团伙的通联模式．

在此基础上，根据不同的分析需求，用堆栈图展示犯罪嫌疑人或者犯罪团伙的通讯频率，以及犯罪事件发生前后 15 天内各个组织的通讯情况．

除了用堆栈图展示通讯流数据特征以外，风筝图也可以表达时间序列数据集的特征．然而，由于风筝图适用于比较变量和属性的“总”值，在涉及跨多个图表比较精确值的详细分析方面比较薄弱．文献[12]在此基础上设计了 Data Vast 花瓶，以展示多变量犯罪数据，可直观、清晰地展示多变量数据在时间上的变化情况，如图 2b 所示．其基于花瓶的图形表示犯罪率在时间上的累加，有助于辅助警务人员分析犯罪数据的时序波动性特征，从整体上把握犯罪事件的规律．

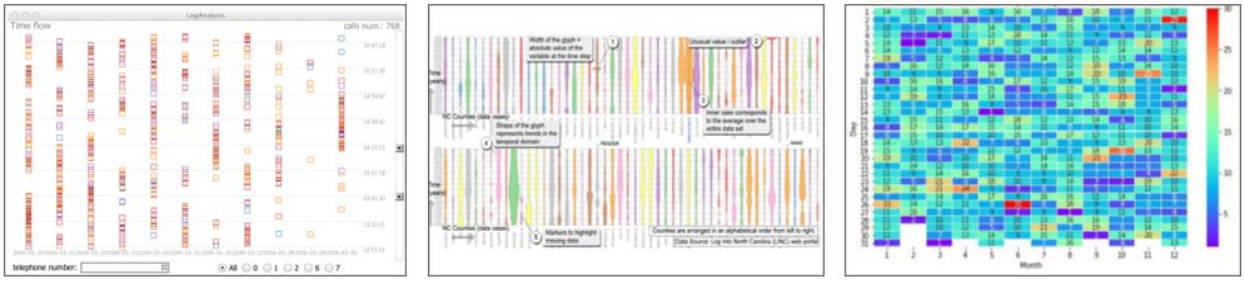
此外，混淆矩阵也常被用于犯罪事件时序分布的数据分析中，例如，文献[5]绘制了伦敦 2016—2018 年平均每月、每天、每小时发生犯罪事件的热力图，对犯罪发生数量和时间做了关联，有效地展示了不同的犯罪模式和嫌疑人活动的时间规律，为警务人员有效地部署警力提供参考，如图 2c 所示．

2.1.2 基于空间的犯罪数据可视化

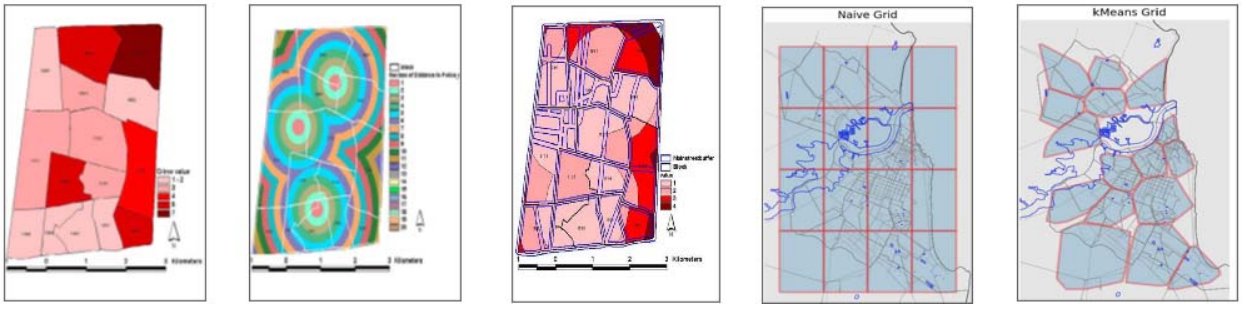
除了具备时序上的分布特征以外，犯罪事件还具有空间上的聚类特性．犯罪数据在犯罪地图上的展示可以根据空间表示法分为 2 大类：点数据(带有犯罪事件的点位置)和区域数据(犯罪事件计数汇总到预先确定的边界)．犯罪点数据可以结合 Focus+Context 交互技术转换或聚合成区域数据[38]．

部分文献采用颜色编码区域内犯罪数据的空间特征[15-17]，以突出犯罪高度集中地区[39]．文献[6]采用基于道路街区的可视化方法描述犯罪聚集特性，如图 2d~图 2f 所示均属于犯罪数据在地图上的规则性映射和空间分析问题[18]．研究人员使用分级统计图显示了基于街区的犯罪率，并从警力覆盖评估和新警局选址推荐 2 个维度，展示了犯罪空间分析的实用价值．

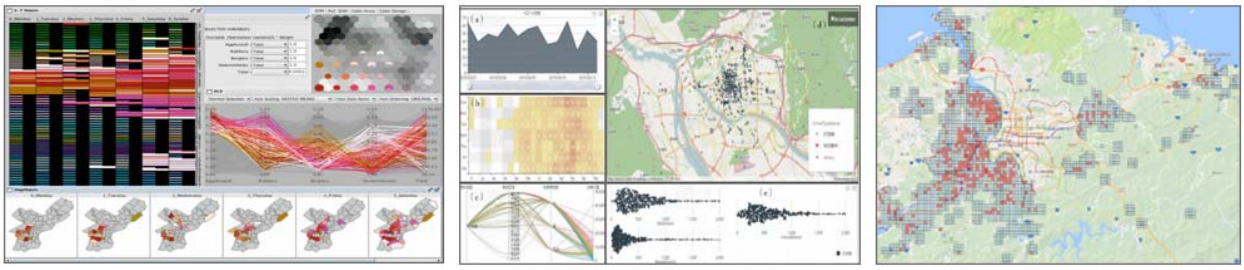
然而，区域内空间属性的变化会使其在描绘基本空间结构时不一致，空间椭圆解决了此问题．例如，最近邻层次聚类基于最近邻点创建热点椭



a. 通讯数据时间流散点图　b. 不同区域犯罪率的时间序列数据花瓶展示　c. 混淆矩阵犯罪事件热力图



d. 街区犯罪率统计　e. 警力覆盖范围　f. 新警局选址　g. 朴素矩形网格　h. *K* 均值聚类网格集



i. 7 个工作日内遭受严重攻击(>25%)的时空趋势　j. 盗窃案件犯罪模式可视分析　k. 毒品相关犯罪热点的可视化地图

图 2　犯罪时空数据可视化

圆. 但是犯罪区域不会是自然椭圆，热点地区的一些地区实际上可能犯罪率较低，同时也忽略了一些重要的空间细节. 不同的是，平滑可以创造连续表面，更容易解释犯罪发生的大致位置.

基于核密度估计(kernel density estimation, KDE)的犯罪热力图可视方法可能会打破行政区划进行犯罪空间描述[19,40]. Chainey 等[17]认为，KDE 是可视化点数据最合适的空间分析技术. de Queiroz Neto 等[41]也将 KDE 聚合得到平方核密度估计(marching squares kernel density estimation, MSKDE)地图进行展示. 除了 KDE 外，有研究者将犯罪热点在空间上表示为网格中的离散单元或凸边形集合，以描述犯罪空间[20]，例如，文献[21]提出了一种基于 *K* 均值聚类的凸多边形集合，可以根据出现坐标调整 *K* 值操纵精确的热点边界，如图 2g 和图 2h 所示.

综上所述，在犯罪数据提炼为犯罪信息的过程中，犯罪地图可视分析方法具有明显的可理解性，在辅助警务人员规划巡逻路线、观察犯罪活动趋势和规律方面具有积极的作用.

目前，在卫星地图全面应用的情形下，通过对城市街景图片进行语义分割，并以不同的颜色进行转化，解读其中蕴含的犯罪信息，可以预测高犯罪率和低犯罪率地区[2-3].

Yin 等[7]和 Lin 等[8]分别利用监督学习和深度学习算法在处理犯罪数据后，将犯罪热点预测映射到地图上，如图 2k 所示. 事实上，将犯罪事件有效地映射到地图上，其背后涉及大量数据挖掘模型、聚类模型的选择和利用等，特别是在地图涉及街道、道路信息时，往往涉及对路网的处理和变形算法. 基于地理信息系统(geographic information system, GIS)的规模聚合网络(scale aggregation network, SANET)工具可以绘制 Voronoi 图来近似市场区域图，但将 Voronoi 图扩展用于网络后的市场区域是确定性的，而实际的市场区域是概率性、边界不明确的[10]. Lau[42]将 3D 地图与犯罪预测相结合，展示了基于密度的聚类算法、网格专题图和密度热图，以识别热点，并展示其与空间特征的关

联，进而得到城市的长期热点.

有研究表明，当数据量较大时，直接将数据按照空间分布映射到地图上，会导致无法从高度重叠的散点图中获取有用信息. 例如，文献[14]采用基于网格代替欧几里得距离的方法，对布鲁塞尔发生的 5144 起交通行为人/工具空间轨迹数据进行分析时出现了上述问题. 这也是缺失时间属性的犯罪空间可视化在犯罪数据深度应用中的局限，尤其是在犯罪预测领域表现得更加突出[43].

2.1.3 基于时空协同的犯罪数据可视化

从犯罪地理学角度上看，只有综合考虑时间和空间 2 个方面的信息，才能完整、全面、准确地揭示犯罪分布特征[44]. 因此，基于探索性的时空协同犯罪可视化在实践中的应用更为广泛.

Qian 等[45]提出了基于网格嵌入的时空关联模型来预测城市中的犯罪数量. 该框架由网格嵌入和犯罪图预测 2 个模块组成 . 其中，网络嵌入模块可以通过考虑地理和上下文空间关系来捕获空间表示向量，并且可以探索空间犯罪基础来表示犯罪图之间的相关性. 犯罪图预测模块可以基于历史空间表示向量来预测下次犯罪图.

文献[46]为犯罪模式的探索性时空分析提供了一个应用框架，将视觉查询工具与聚类分析相结合，其优势是能够将犯罪数据从时间和空间 2 个维度进行聚合，使用时空立方体显示空间扫描统计的结果，但信息维度比较单一.

Guo 等[9]运用时空协同分析犯罪属性的时空分布及不同区域各类犯罪的时间变化趋势，时空多元可视化(visualization for space time and multivariate patterns, VIS-STAMP)将立方体视为排列在空间犯罪矩阵中的一组时间序列，综合运用了系列地图、平行坐标图、自组织地图和时间矩阵图等，并提供了如图 2i 所示探索性交互手段、可视分析犯罪数据的多维度特征.

多视图协同是犯罪时空可视化发展的常用方法. 文献[4]构建了一个基于多视图协同的犯罪数据可视化分析系统，通过如图 2j 所示交互式可视分析可发现基于宏观时空规律的犯罪模式. 文献[47]将犯罪时空分析的视野从宏观分析聚焦到局部地区的犯罪模式探索上，代表了犯罪规律挖掘的常用方法. 同时，也弥补了文献[46]依靠视觉观察、缺乏严格的统计测试程序来评估犯罪严重程度的缺陷.

然而，在微观层面，语义信息的缺乏难以将犯罪轨迹与嫌疑人的活动联系起来. 文献[23]利用甘特图将嫌疑人在连续时间段中的各种活动序列化，构建了一个由多视图组成的交互式可视分析系统，从时间、空间、人和语义的角度对嫌疑人进行轨迹分析，为案件侦查和法庭举证提供帮助.

总体来说，犯罪时空数据的可视分析不仅能够发现历史犯罪事件的移动规律性，还能够探索嫌疑人的行为轨迹，进而分析犯罪演变的模式.

近年来，基于时空维度的犯罪预测研究逐步向犯罪时空数据和当地社会、人口、经济发展等相关因素相结合的综合分析模型方向发展，这体现了犯罪分析研究的新特征.

2.2 犯罪文本数据可视化

犯罪文本的来源通常是案件文档和法律文书，文本分析需要先将其进行分词和信息抽取，再进行词频统计或者相似度计算. 词频统计有助于分析犯罪热点和案件研判，文本相似度可视化可为案件的串并分析提供支持，但仅依靠人工智能不能精准对分析文本数据的细粒度，仍需要依赖于专家的先验知识.

2.2.1 基于词云的犯罪文本信息展现

关键词云图作为文本可视化方法中一种最常见和具有视觉吸引力的方法，将文本精炼为一些高频出现的词，能较好地完成犯罪文本信息的展现. 文献[24]将 Worldle 图运用到了案情文本数据. Holzinger 等[48]提出了图 3a 所示可视分析框架，以探索文本证据与计算机取证的相对关系.

2.2.2 基于数据挖掘的犯罪文本可视分析

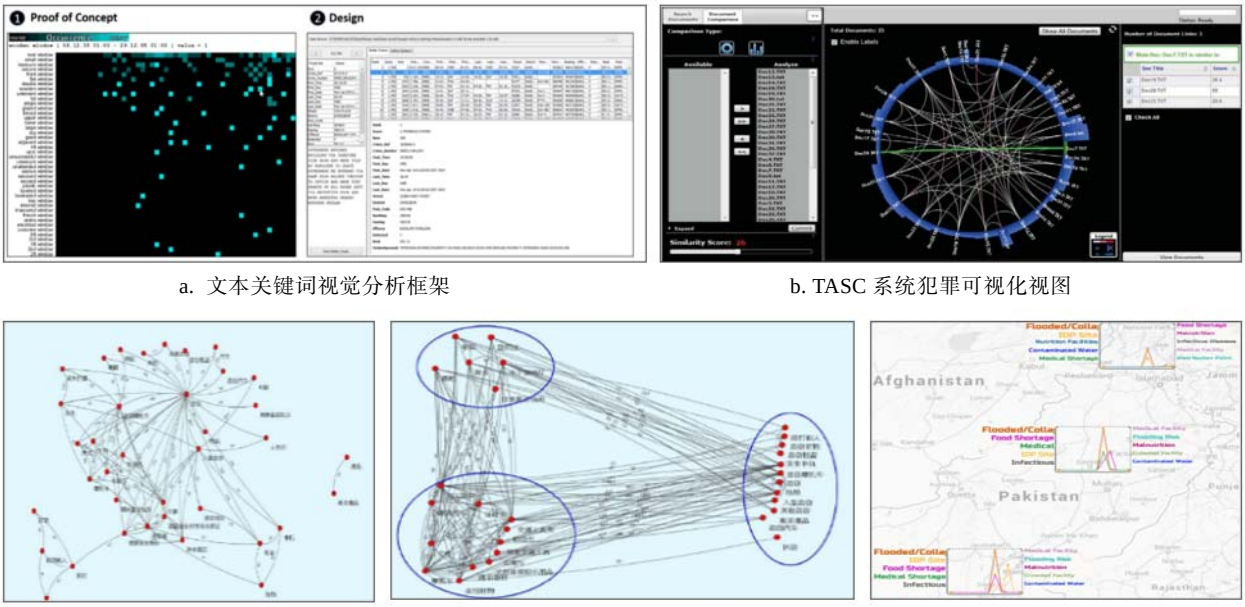
词频分析结合词云图只能简单地展示犯罪文本关键词的统计特征. 为了深入探索犯罪词汇之间的相关性，以及犯罪文本数据隐含的丰富信息，需要结合数据挖掘技术和关联视图才能实现较好分析效果. 文本聚类一般使用 *K*-means 算法，李强[13]优化了 *K*-means 算法，提高了聚类中心的凝聚力，并利用知识图谱展现，如图 3c 和图 3d 所示. 第 1 层知识图谱展示中心主体，第 2 层知识图谱反映案事件的规律性和犯罪的某些行为模式.

Ku 等[49]开发了一个同类犯罪文本分析系统(textual analysis of similar crimes, TASC)，利用文档相似性算法来分析犯罪报告之间的语义特征. 在技术角度层面，该系统更倾向于高水平的可视化和低水平的文本突出显示相结合；在应用角度层面，该系统提供了不同的数据视图，可提高对犯罪案件相似性比对分析的能力，如图 3b 所示.

2.2.3 基于多源数据融合的犯罪文本可视分析

根据犯罪事件的多源相关性，为提高对犯罪

事件的认知和分析能力，有必要将犯罪文本数据融合其他数据进行可视分析．例如，文献[25]将文本与空间相结合，使用图 3e 所示地理时间标签可视分析(geo-temporal tag visualization, GTViz)方法将文本呈现在空间地图上，用于观察该地区一段时期以来关键事件的发生情况，从而帮助政府解决危机问题，但其可视化手段和水平仍需进一步丰富和提高．



a. 文本关键词视觉分析框架　　b. TASC 系统犯罪可视化视图

c. 第 1 层知识图谱　　d. 第 2 层知识图谱　　e. 巴基斯坦的地理时间标签视图

图 3　犯罪文本数据可视化

2.3　犯罪网络关系数据可视化

警务工作中关于人员、案件间的关系数据十分丰富，挖掘与分析人员、案件之间的关系特征，构建人员、案件之间的关系网络，对于案情分析、有组织犯罪侦查、重点人员预警等工作具有十分重要的作用．本节分别从基于图论的犯罪网络关系生成、犯罪网络关系静态可视分析、犯罪网络关系动态演化可视分析及基于多源数据融合犯罪网络生成与可视分析的角度进行归纳．

2.3.1　基于图论的犯罪网络关系生成

将犯罪数据进行网络可视化建模时，最直接的可视编码方法是节点代表个体，线代表关系，线条的粗细表示关系的强弱[50]．许多学者使用网络分析工具来揭示犯罪实体之间的关系．例如，Klerks[51]将现有的犯罪网络关系分析方法和工具分为 3 代：第 1 代是手动方法，十分低效；第 2 代可以自动生成犯罪网络关系的图形；第3代是基于系统网络体系结构的可视化方法．大多数现有的犯罪网络关系分析方法和工具均属于第 2 代．其中，Analyst’s Notebook, Netmap 和 XANALYS Link Explorer 最受欢迎．

以 Netmap[26]为例，第 2 代方法提供了网络可视化功能．如图 4a 所示该系统在圆的周长上布置实体，用直线来表示链接，据此发现隐藏在网络背后的交互和联系．

虽然第 2 代工具能够使用多种可视编码来可视化犯罪网络，但其缺乏分析功能，难以挖掘数据隐含的模式，故仍然需要依赖分析人员深入研究图表，从中发现网络的结构属性．

2.3.2　犯罪网络关系静态可视分析

为了减少人工劳动，更便捷地分析犯罪网络关系．系统网络体系结构(systems network architecture, SNA)作为第 3 代分析工具被广泛应用，能够提供更先进的分析功能和复杂的结构分析工具(visual analysis of financial activity networks, VisFAN)，可便捷地发现关于犯罪网络结构和组织的有用知识[27-28]．

随着手机智能化应用的普及，社交网络数据在犯罪分析中具有越来越重要的作用．例如，White 等[29]提供了一款应用 TwitterHitter，利用地理信息科学和可视化技术，从社交网络特别是 Twitter 网站中获取犯罪嫌疑人的时空信息和通信关系网，用于绘制犯罪嫌疑人在 Twitter 近期活动的地图和有针对性的犯罪网络图，帮助打击整个

犯罪团伙. Ferrara 等[37]对手机通信网络中的犯罪网络关系进行了分析与检测，从电话记录中重建网络和探索犯罪组织，如图 4b 所示.

Qazi 等[52]提出了基于图论的犯罪匹配行为与时间-空间知识图谱. 犯罪匹配过程通常涉及冗长且信息密集的任务，即在犯罪行为人之间引发似是而非的关联以识别潜在嫌疑人. 针对这一过程的辅助，该研究展示了联想搜索的应用，用于从信息中唤起可能的关联，以促进犯罪匹配过程. 其中，树结构和图标图形组成的知识图谱发挥了关键作用.

针对网络金融犯罪，VisForFraud 系统[53]可以自动检索结构化数据，描述实体对之间的静态关系类型.

2.3.3 犯罪网络关系动态演化可视分析

犯罪网络关系的静态分析在揭示犯罪网络的结构和性能方面取得了巨大成果，但实践中更多的犯罪网络关系是随时间发生变化的. Dynalink[54]将高维数据可视化与时间动态结合，随着时间的推移、互助网络的动态发展和数据更新，可视化效果可实时更新.

此外，网络攻击的日志数据以及溯源数据也具有明显的时间和网络属性. 为了避免此类攻击，网络攻击数据动态可视分析便于发现攻击来源，已应用于各级安全防护，如防火墙数据分析[30]、入侵检测系统[31-32]等.

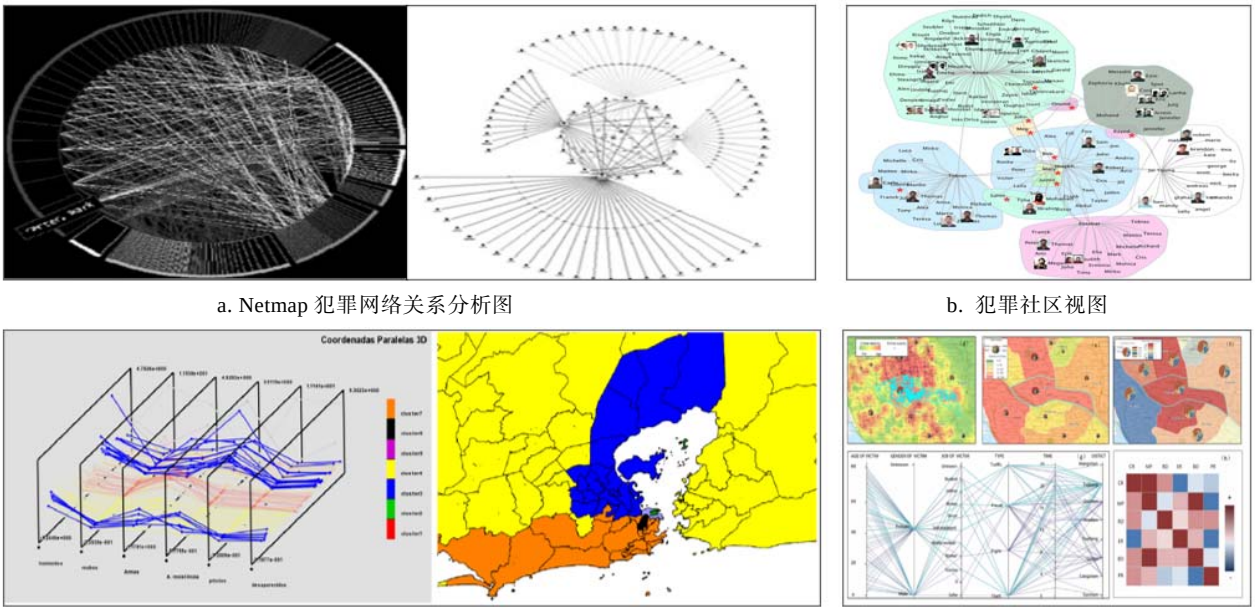
2.3.4 基于多源数据融合犯罪网络生成与可视分析

随着公共安全感知体系不断完善和标准化，更多的犯罪网络关系数据呈现多维态势. 分析数据的多维属性，有助于揭示犯罪数据各属性之间的关系，反映犯罪数量与所处的区域、社会、经济、人口和环境之间的相互作用关系，有效地辅助犯罪模式分析、影响因素分析及警力资源配置[11].

Sellam 等[55]设计了 Claude 系统，利用维度和度量之间的非线性统计相关性来进行大规模数据可视化，展示社会因素、家庭情况、经济地位等因素对犯罪的影响. 该模式的优势是可以全面检查每个可能的维度组合与犯罪的相关性，快速发现犯罪率与相关因素之间的特殊关系，其在可视化方面的缺陷可以利用数据投影(如 Scagnostics)进行弥补.

Syamkumar 等[56]引入了 Bigfoot 可视化系统，用于描述边界网关协议(border gateway protocol, BGP)更新的情况，用于突出和评估网络更新流中的各种行为模式. 通过 IP 地理定位和二维多边形的可视化，进行网络足迹多边形的不同表示. 最后，通过集中可视化 139 个事件，展示了网络犯罪中的犯罪足迹.

从高维数据中提取有价值的信息是一个具有挑战性的问题[11]. 其中一个挑战是包含相关信息可能性表示的数量随数据量和维度增加呈指数增长. Behrisch 等[57]利用聚焦上下文和语义缩放接口



a. Netmap 犯罪网络关系分析图　b. 犯罪社区视图

c. 3D 平行坐标与空间关系视图　d. 多元可视化展现

图 4 犯罪网络关系数据可视化

的方法引导用户到感兴趣的区域，以避免信息噪声干扰.

利用平行坐标图表现犯罪网络关系多维属性数据在各维度上的分布是较为常见的方法. 例如，文献[33]采用平行坐标图展示犯罪数据多维属性关联数量；文献[58]使用刷涂和链接技术，以立体平行坐标轴可视化多维数据，如图 4c 所示.

为了考虑犯罪事件的潜在影响或依赖关系，除了平行坐标轴外，还可以将犯罪网络关系数据和时空数据相结合，在地图上展示犯罪热点，设计可视图元来展示犯罪数据的区域特征，使用混淆矩阵表示数据相关性的强度，实现跨视图的方式进行展示[22,59]，如图 4d 所示，这也是近年来比较流行的犯罪数据可视化展示方式.

3　总结与展望

本文对基于警务工作场景的犯罪数据可视化任务进行了归纳和梳理，并从时空数据、文本数据和网络数据等维度对犯罪数据可视化方法和技术手段进行了调研和综述. 这些工作在辅助警务部门打击治理犯罪、提升警务管理和执法效能等方面发挥了重要作用，但面对大数据时代各类犯罪数据呈指数级增长，犯罪数据分析应用场景越来越复杂的态势，传统的可视化方法和技术需要不断迭代升级才能满足分析复杂治安环境的需求，主要表现在以下几个方面.

(1) 犯罪模式发现对犯罪知识洞察的时效性和可视方式多样性要求越来越高. 如何找到更快速训练的深度学习方法，保证犯罪核心要素(人、案、地、事、物)知识图谱嵌入查询高精准度的同时，提高实时可视响应、增强现实-虚拟现实等新技术的多维度虚拟现实可视分析算法研究等均是比较有价值的研究方向.

(2) 警力资源配置的科学性依赖于犯罪时空可视化的交互性和犯罪文本可视化的地理信息特性. 比较有实战价值的研究方向是利用反馈驱动的视图探索框架对专题视图进行过滤，在文本地图中对定性数据的定量化展示，以及与地理位置相关联的警务化标签进行可视化与信息检索，以支撑警务资源调度和勤务体制改革.

(3) 案件侦查研判需要有精准、高效、全面的犯罪网络关系分析. 随着警务部门采集的犯罪数据内容更加丰富、维度更加多样、频率更加实时，越来越多的犯罪数据同时包含时空、网络、文本、语音、语义等属性，针对多数据类型融合的协同分析与计算可视化跨平台展示，也将成为犯罪数据可视化发展面临的一大挑战.

此外，由于当前我国犯罪数据的公开面临一些问题，导致国内犯罪学领域实证研究的匮乏，在某种程度上，掣肘了大数据时代下犯罪数据可视化技术的进步和犯罪数据价值的释放. 为了促进犯罪数据可视化研究，需要探讨更加多样、更加全面的犯罪统计数据的科学公开机制，研究建立交互式的犯罪数据公开平台的可行性[60]，切实发挥犯罪数据分析在犯罪学研究和犯罪防控，乃至社会综合治理中日益突出的支撑作用.

参考文献(References):

[1] Gorr W, Oligochaetes A, Thompson Y. Short-term forecasting of crime[J]. International Journal of Forecasting, 2003, 19(4): 579-594

[2] Amiruzzaman M, Curtis A, Zhao Y, *et al*. Classifying crime places by neighborhood visual appearance and police geonarratives: a machine learning approach[J]. Journal of Computational Social Science, 2021, 4(2): 813-837

[3] Kadiyam, Praneeth. Crime rate prediction from street view images using convolutional neural networks and transfer learning[D]. Enschede: University of Twente, 2021

[4] Liu Mingchao, Wu Sheng, Yu Jinsongdi. Spatio-temporal data visual analysis method based on multiple-view collaboration[J]. Journal of Geomatics Science and Technology, 2017, 34(2): 211-214+220(in Chinese)
(刘明超，吴升，余劲松弟. 基于多视图协同的时空数据可视分析方法[J]. 测绘科学技术学报，2017，34(2)：211-214+220)

[5] He Wei. Visual analysis of police data[J]. Journal of Fujian Computer, 2020, 36(6): 81-83(in Chinese)
(何巍. 警务数据可视化分析[J]. 福建电脑，2020，36(6)：81-83)

[6] Leitner M. Crime modeling and mapping using geospatial technologies[M]. Heidelberg: Springer, 2013: 1-446

[7] Yin J R, Michael I A, Afa I J. Machine learning algorithms for visualization and prediction modeling of Boston crime data[OL]. [2022-01-06]. https://www.preprints.org/manuscript/202002.0108/v1

[8] Lin Y L, Chen T Y, Yu L C. Using machine learning to assist crime prevention[C] //Proceedings of the 6th IIAI International Congress on Advanced Applied Informatics. Los Alamitos: IEEE Computer Society Press, 2017: 1029-1030

[9] Guo D S, Wu J. Understanding spatiotemporal patterns of multiple crime types with a geovisual analytics approach[M]. Heidelberg: Springer, 2013: 367-385

[10] Okabe A, Okunuki K I, Shiode S. The SANET toolbox: new methods for network spatial analysis[J]. Transactions in GIS, 2006, 10(4): 535-550

[11] Li Daichao, Wu Sheng. Theme-oriented visual analysis of

crime with big data[J]. Journal of Geo-Information Science, 2014, 16(5): 735-745(in Chinese)
(李代超, 吴升. 面向不同主题的犯罪大数据可视分析[J]. 地球信息科学学报, 2014, 16(5): 735-745)

[12] Thakur S, Rhyne T M. Data vases: 2D and 3D plots for visualizing multiple time series[C] //Proceedings of International Symposium on Visual Computing. Heidelberg: Springer, 2009: 929-938

[13] Li Qiang. Research and implementation on intelligence analysis system based on content analysis[D]. Changsha: National University of Defense Technology, 2012(in Chinese)
(李强. 基于内容分析法的情报分析系统研究与实现[D]. 长沙: 国防科学技术大学, 2012)

[14] Aerts K, Lathuy C, Steenberghen T, *et al*. Spatial clustering of traffic accidents using distances along the network[OL]. [2022-01-06].https://www.ictct.net/wp-content/uploads/19-Minsk-2006/19-Aetts-Full-paper.pdf

[15] ToppiReddy H K R, Saini B, Mahajan G. Crime prediction & monitoring framework based on spatial analysis[J]. Procedia Computer Science, 2018, 132: 696-705

[16] Heim K. Visualization and modeling for crime data indexed by road segments[D]. Fairfax: George Mason University, 2014

[17] Chainey S, Tompson L, Uhlig S. The utility of hotspot mapping for predicting spatial patterns of crime[J]. Security Journal, 2008, 21(1): 4-28

[18] Townsley M. Crime mapping and spatial analysis[C] //LeClerc B, Savona E. Crime Prevention in the 21st Century. Heidelberg: Springer, 2017: 101-112

[19] Morgan J D. A visual time-geographic approach to crime mapping[D]. Florida: The Florida State University, 2010

[20] Araújo A, Cacho N, Bezerra L, *et al*. Towards a crime hotspot detection framework for patrol planning[C] //Proceedings of the 20th IEEE International Conference on High Performance Computing and Communications; the 16th IEEE International Conference on Smart City; the 4th IEEE International Conference on Data Science and Systems. Los Alamitos: IEEE Computer Society Press, 2018: 1256-1263

[21] Borges J, Ziehr D, Beigl M, *et al*. Feature engineering for crime hotspot detection[C] //Proceedings of the IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation. Los Alamitos: IEEE Computer Society Press, 2017: 1-8

[22] Morshed A, Tsai P W, Jayaraman P P, *et al*. VisCrime: a crime visualisation system for crime trajectory from multi-dimensional sources[C] //Proceedings of the 12th ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2019: 802-805

[23] Agarwal J, Nagpal R, Sehgal R. Crime analysis using K-means clustering[J]. International Journal of Computer Applications, 2013, 83(4): 1-4

[24] Li Daichao. A study on criminal visualization[D]. Fuzhou: Fuzhou University, 2014(in Chinese)
(李代超. 犯罪信息可视化研究[D]. 福州: 福州大学, 2014)

[25] Aman H, Irani P, Amini F. Revisiting crisis maps with geo-temporal tag visualization[C] //Proceedings of the IEEE Pacific Visualization Symposium. Los Alamitos: IEEE Computer Society Press, 2014: 177-184

[26] Aghakhani S, Dawoud K, Alhajj R, *et al*. A global measure for estimating the degree of organization and effectiveness of individual actors with application to terrorist networks[M]. Heidelberg: Springer Press, 2011: 189-222

[27] Xu J, Chen H. Criminal network analysis and visualization[J]. Communications of the ACM, 2005, 48(6): 100-107

[28] Didimo W, Liotta G, Montecchiani F. Network visualization for financial crime detection[J]. Journal of Visual Languages & Computing, 2014, 25(4): 433-451

[29] White J J D, Roth R E. TwitterHitter: geovisual analytics for harvesting insight from volunteered geographic information[OL]. [2022-01-06]. https://www.researchgate.net/publication/228375652

[30] Golnabi K, Min R K, Khan L, *et al*. Analysis of firewall policy rules using data mining techniques[C] //Proceedings of the IEEE/IFIP Network Operations and Management Symposium NOMS. Los Alamitos: IEEE Computer Society Press, 2006: 305-315

[31] Yelizarov A, Gamayunov D. Visualization of complex attacks and state of attacked network[C] //Proceedings of the 6th International Workshop on Visualization for Cyber Security. Los Alamitos: IEEE Computer Society Press, 2009: 1-9

[32] Xu Wei. Design and implementation of real-time visualization system for network intrusion detection[D]. Beijing: Beijing University of Posts and Telecommunications, 2020(in Chinese)
(徐维. 网络入侵检测的实时可视化系统设计与实现[D]. 北京: 北京邮电大学, 2020)

[33] Silva L J S, González S F, Almeida C, *et al*. CrimeVis: an interactive visualization system for analyzing crime data in the state of Rio de Janeiro[C] //Proceedings of the 19th International Conference on Enterprise Information Systems. Porto: SciTe Press, 2017: 193-200

[34] Eck J, Chainey S, Cameron J, *et al*. Mapping crime: understanding hotspots[M]. Washington: National Institute of Justice Press, 2005: 1-71

[35] Chainey S, Ratcliffe J. GIS and crime mapping[M]. London: John Wiley & Sons Press, 2013: 1-45

[36] Zheng Ziwan, Jin Cheng, Eck J E, *et al*. Crime mapping: understanding crime hotspots (middle)[J]. Juvenile Delinquency Prevention Research, 2012(5): 67-77(in Chinese)
(郑滋椀, 金诚, Eck J E, 等. 犯罪制图: 理解犯罪热点(中)[J]. 预防青少年犯罪研究, 2012(5): 67-77)

[37] Ferrara E, de Meo P, Catanese S, *et al*. Detecting criminal organizations in mobile phone networks[J]. Expert Systems with Applications, 2014, 41(13): 5733-5750

[38] Wang Z L, Liu L, Zhou H L, *et al*. How is the confidentiality of crime locations affected by parameters in kernel density estimation?[J]. ISPRS International Journal of Geo-Information, 2019, 8(12): 544

[39] Mohler G O, Short M B, Brantingham P J, *et al*. Self-exciting point process modeling of crime[J]. Journal of the American Statistical Association, 2011, 106(493): 100-108

[40] Tompson L. Crime mapping/geospatial information systems[C] //The Encyclopedia of Research Methods in Criminology and Criminal Justice. London: John Wiley & Sons Press, 2021: 20-27

[41] de Queiroz Neto J F, Santos E, Vidal C A, *et al*. A visual analytics approach to facilitate crime hotspot analysis[J]. Computer Graphics Forum, 2020, 39(3): 139-151

[42] Lau H Y. Discovering spatio-temporal pattern of city crime-visual analysis on felony crime in New York[OL]. [2022-01-06]. https://www.preprints.org/manuscript/202101.0292/v1

[43] Liu H, Brown D E. Criminal incident prediction using a point-pattern-based density model[J]. International Journal of Forecasting, 2003, 19(4): 603-622

[44] Cheng Wei, Wu Jianping. A research review on spatio-temporal distribution of crime[J]. World Regional Studies, 2013, 22(2): 151-158(in Chinese)
(程薇, 吴健平. 国外犯罪时空分布研究综述[J]. 世界地理研究, 2013, 22(2): 151-158)

[45] Qian Y T, Pan L, Wu P, *et al*. GeST: a grid embedding based spatio-temporal correlation model for crime prediction[C] //Proceedings of the 5th IEEE International Conference on Data Science in Cyberspace. Los Alamitos: IEEE Computer Society Press, 2020: 1-7

[46] Cheng T, Williams D. Space-time analysis of crime patterns in central London[J]. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012, XXXIX-B2: 47-52

[47] Velazquez M V, Kounadi O, Podor A. Analysis and mapping of crime perception: a quantitative approach of sketch maps[J]. AGILE: GIScience Series, 2020, 1: 1-18

[48] Holzinger A, Pasi G. Human-computer interaction and knowledge discovery in complex, unstructured, big data[M]. Heidelberg: Springer Press, 2013: 160-172

[49] Ku C H, Nguyen J H, Leroy G. TASC-crime report visualization for investigative analysis: a case study[C] //Proceedings of the IEEE International Conference on Information Reuse & Integration. Los Alamitos: IEEE Computer Society Press, 2012: 466-473

[50] Xu J, Marshall B, Kaza S, *et al*. Analyzing and visualizing criminal network dynamics: a case study[C] //Proceedings of Intelligence and Security Informatics. Heidelberg: Springer, 2004: 359-377

[51] Klerks P. The network paradigm applied to criminal organisations: theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands[C] //Klerks P. Transnational organised crime. London: Routledge, 2003: 111-127

[52] Qazi N, Wong B L W. Behavioural & tempo-spatial knowledge graph for crime matching through graph theory[C] //Proceedings of European Intelligence and Security Informatics Conference. Los Alamitos: IEEE Computer Society Press, 2017: 143-146

[53] Giacomo E D, Didimo W, Liotta G, *et al*. Visual analysis of financial crimes[C] //Proceedings of the International Conference on Advanced Visual Interfaces. New York: ACM Press, 2010: 393-394

[54] Park A J, Tsang H H, Brantingham P L. Dynalink: a framework for dynamic criminal network visualization[C] //Proceedings of the European Intelligence and Security Informatics Conference. Los Alamitos: IEEE Computer Society Press, 2012: 217-224

[55] Sellam T, Müller E, Kersten M. Semi-automated exploration of data warehouses[OL]. [2022-01-06]. https://dl.acm.org/doi/abs/10.1145/2806416.2806538

[56] Syamkumar M, Durairajan R, Barford P. Bigfoot: a geo-based visualization methodology for detecting BGP threats[C] //Proceedings of the IEEE Symposium on Visualization for Cyber Security. Los Alamitos: IEEE Computer Society Press, 2016: 1-8

[57] Behrisch M, Korkmaz F, Shao L, *et al*. Feedback-driven interactive exploration of large multidimensional data supported by visual classifier[C] //Proceedings of the IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2014: 43-52

[58] Li D C, Wang Y J, Wu S, *et al*. An visual analytics approach to explore criminal patterns based on multidimensional data[C] //Proceedings of the IEEE International Geoscience and Remote Sensing Symposium. Los Alamitos: IEEE Computer Society Press, 2017: 5563-5566

[59] Xiao J, Zhou X L. Crime exposure along my way home: estimating crime risk along personal trajectory by visual analytics[J]. Geographical Analysis, 2020, 52(1): 49-68

[60] Ruan Chongjun. A study on crime data disclosure in China in the era of big data—from the perspective of comparative study between China and America[J]. Journal of Shandong Police College, 2019, 31(4): 109-118(in Chinese)
(阮重骏. 大数据时代下我国犯罪数据公开研究——以中美比较研究为视角[J]. 山东警察学院学报, 2019, 31(4): 109-118)