

时空分析和 K 近邻算法在犯罪分析中的应用研究

王娟¹ 兰月新² 吴春颖¹ 陈蕾¹ 张双狮¹

¹(中国人民警察大学部队管理系 河北 廊坊 065000)

²(中国人民警察大学边防系 河北 廊坊 065000)

摘 要 犯罪是当今世界面临的普遍社会问题,准确实时的犯罪预测有助于预防犯罪、降低犯罪率。但由于犯罪事件之间的关联以及与周围环境的复杂关系,犯罪分析与预测至今仍然是一个具有挑战性的问题。本文采用了基于时空分析的可视化技术和 K 近邻算法来分析不同犯罪类型在伦敦区域内的分布。首先,对犯罪数据进行预处理和可视化;然后,通过 K 近邻算法对犯罪数据集进行分析、提取有用信息,并通过可视化的方式展现数据之间的关联关系;最后,通过对不同类型犯罪进行分析,预测不同犯罪类型的高发区域和高发时间。因此有助于公安执法人员的警力部署和指挥,从而达到预防犯罪的目的。

关键词 位置预测;时空特征;K 近邻算法;犯罪预测;可视化

中图分类号 TP39 DOI:10.16707/j.cnki.fjpc.2019.07.009

Research on the Application of Spatial-Temporal Analysis and KNN Algorithm in Crime Analysis

WANG Juan¹, LAN Yuexin², WU Chunying¹, CHEN Lei¹, ZHANG Shuangshi¹

¹(Force Management Department, China People's Police University, Langfang, China, 065000)

²(Border Department, China People's Police University, Langfang, China, 065000)

1 引言

犯罪是当今世界面临的普遍社会问题,它不仅影响着人们的生活、经济增长,而且影响着人们的社会安全感,是社会安全领域不可忽视的问题。为了降低犯罪率,预测预警犯罪的发生,公安执法人员需要借助大数据分析手段来预测犯罪和分析犯罪。准确的实时的犯罪预测有助于预防犯罪、降低犯罪率,但由于犯罪事件之间的关联以及与周围环境的复杂关系,犯罪分析与预测至今仍然是一个具有挑战性的问题。

Hitesh K.通过多种可视化技术和机器学习算法来预测犯罪在一个区域的分布^[1]。为了优化犯罪事件的分析和预测,还提出了一种基于空间分析的监测框架,用于可视化犯罪网络,并通过使用谷歌地

图和各种软件包的机器学习算法对其进行分析。Kaiming He 提出了一个残差学习框架,以便对网络深度更深的网络进行培训^[2]。残差学习框架在深度高达 152 层的残差网络实现了 3.57% 的误差率,但仍然保持较低的复杂度。Tao Hu 建立了一个城市犯罪的时空贝叶斯模型,用于分析时空模式并确定发展趋势^[3]。并将该模型用于武汉市犯罪数据,发现入室盗窃犯罪率与每个社区的平均常住人口数量和当地网吧数量正相关。王增利提出了一种基于时空影响范围的网络构造方法,构造了一种基于节点影响强度的犯罪传输网络,并引入复杂网络的度、平均度、聚集系数等特征参数分析犯罪传输网络^[4]。研究结果表明近邻的时空单元的犯罪率具有一定的关联关系,犯罪聚集系数大小与未来犯罪率的变化具有关联性。段炼提出了融合时空语义的位置时

本文得到河北省统计科研计划项目(No.2018HZ03)、河北省高等教育教学改革研究与实践项目(No.2018GJJG450)、河北省重点研发计划项目(No.18210806)资助。王娟(通信作者),女,1979年生,主要研究领域为大数据技术、计算机视觉。E-mail:66267476@qq.com。兰月新,男,1981年生,主要研究领域为网络舆情。E-mail: lanyuexin@wjxy.edu.cn。吴春颖,女,1977年生,主要研究领域为人工智能。E-mail: 775251252@qq.com。陈蕾,女,1979年生,主要研究领域为大数据技术。E-mail: chenlei@wjxy.edu.cn。张双狮,男,1974年生,主要研究领域为人工智能。E-mail: shuangshizhang@qq.com。

空预测模型 SSLP, 利用疑犯在不同语义时段和语义位置上的分布邻近性提取目标疑犯的相似疑犯群体, 并采用贝叶斯模型实现疑犯的时空预测^[5]。

2 K 近邻算法

K 近邻 (K-Nearest Neighbor, KNN) 算法是有监督学习方法之一。当对测试样本进行分类时, 首先扫描训练样本集, 找到与该样本集最相似的 k 个样本, 根据 k 个样本的类别进行确定测试样本的类别。该算法对异常数据不敏感, 简单易实现, 在训练数据集较大时, 效果较好。对某点的属性的判断以其相邻点的多数票进行分类, 该算法也适用于犯罪数据集。即相同犯罪类型的案件在犯罪时间、地点上具有重复性, 前一个犯罪地点附近的地区更有可能发生犯罪。因此, 地点、时间、犯罪类型是需要考虑的重要因素。

2.1 马氏距离计算方法

马氏距离 (Mahalanobis Distance), 与欧氏距离和曼哈顿距离不同, 马氏距离考虑了不同变量之间的相关关系。在计算两个样本的距离时, 不仅考虑样本本身的信息, 且兼顾样本所在样本集的信息 (即协方差矩阵)。而且变量的取值范围不会对样本距离计算产生影响, 可以直接应用在没有进行标准化的数据中。对于本文应用的犯罪数据集, 以经度、纬度、时间等作为维度。假设多维空间的维度为 d , $x_1 = (x_{11}, x_{12}, \dots, x_{1d})$, $x_2 = (x_{21}, x_{22}, \dots, x_{2d})$ 是两个样本点, 那么 x_1 和 x_2 的马氏距离为:

$$d(x_1, x_2) = \sqrt{(x_1 - x_2)^T \Sigma^{-1} (x_1 - x_2)} \quad (1)$$

2.2 k 值的选取

K 近邻算法对训练样本进行分类时, k 值的选取会影响模型的准确度。因此一般采取交叉验证法, 评估一系列不同的 k 值, 选取结果最好的 k 值作为训练参数。然后, 从训练样本中得到 k 个与测试样本邻近的最相似的样本, 根据 k 个最相似的训练样本的类别, 通过邻居投票的方式来确定测试样本的类别。

3 实验与分析

本文所使用的数据集是从英国警察局的官方网站上获得的可靠、真实数据。该数据集共包含 11 个属性, 其中 5 个属性即犯罪类型、位置、日期、纬度和经度被用于犯罪分析研究。2017 年和 2018

年伦敦的犯罪数据集分别作为训练数据集做对比分析研究。在预处理阶段, 删除不一致的数据 (如缺失值、异常值等) 并对数据进行正则化、归一化处理。

图 1 和图 2 分别为 2017 年、2018 年伦敦地区月度犯罪趋势分析图。从图中可以看出, 2017 年犯罪数量居前三的犯罪类型是反社会行为、其他盗窃、暴力和性犯罪分别占到了该区域全年总犯罪数量的 23.2%、21.0% 和 12.5%, 即前三大犯罪类型占到了总犯罪数量的 56.5%。对比 2018 年, 排名前三的犯罪类型是其他盗窃、反社会行为、暴力和性犯罪分别占到了该区域全年总犯罪数量的 20.5%、18.2%、14.6%, 即前三大犯罪类型占到了总犯罪数量的 52.5%。通过数据可以看出, 伦敦地区 2018 年较 2017 年, 其他盗窃犯罪行为从 2017 年的 1451 例上升到 2018 年的 1628 例, 成为 2018 年全年犯罪数量最多的犯罪类型。而反社会行为犯罪从 2017 年的 1600 例下降到 2018 年的 1442 例, 数量有明显的下降。

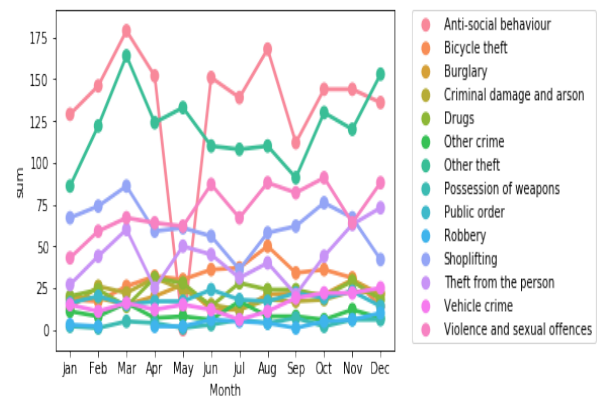


图 1 2017 年伦敦地区月度犯罪趋势分析图

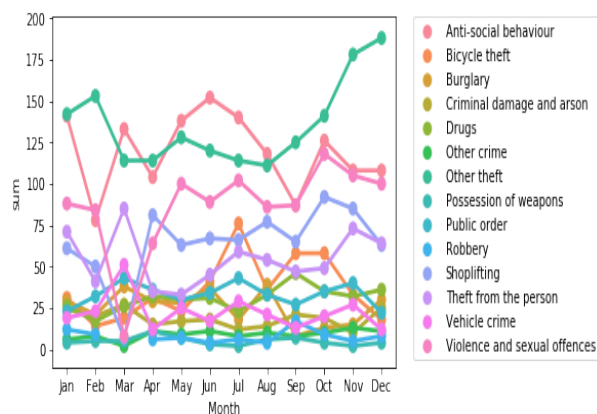


图 2 2018 年伦敦地区月度犯罪趋势分析图

图 3、图 4 分别是 2017 年、2018 年伦敦地区犯罪类型时空分布图。从图中，我们可以看到 14 种不同犯罪类型的地点分布。

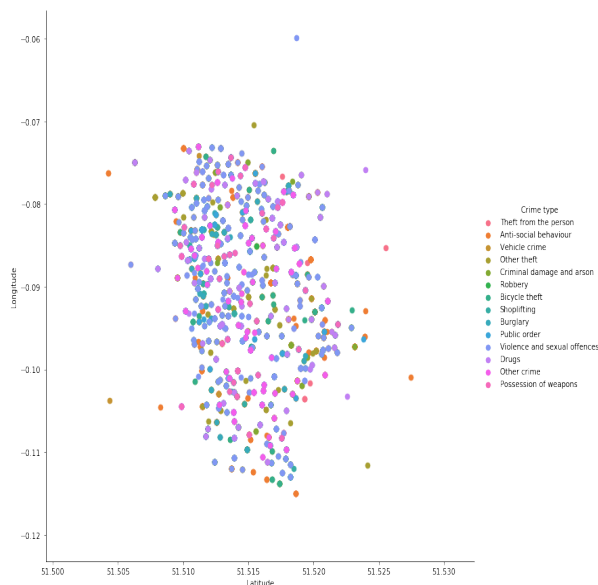


图 3 2017 年伦敦地区犯罪类型时空分布图

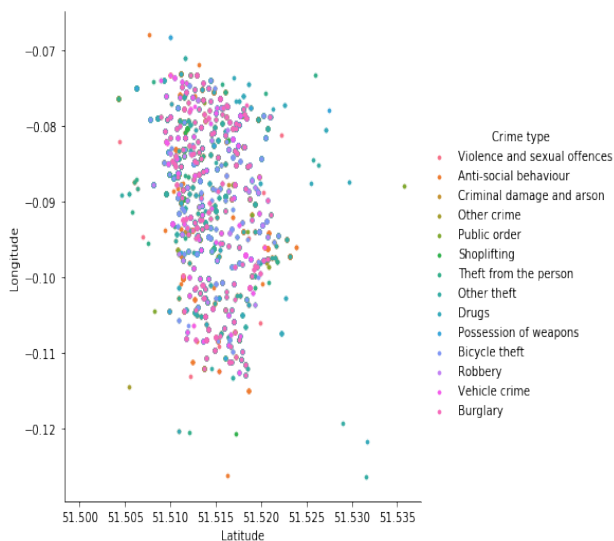


图 4 2018 年伦敦地区犯罪类型时空分布图

4 结束语

本文采用了多种可视化技术和 K 近邻算法来分析不同犯罪类型在伦敦区域的时空分布。首先，对犯罪数据进行预处理和可视化；然后，利用 K 近邻算法从犯罪数据集中提取有效信息，并通过可视化的方式展现数据之间的关联关系；最后，通过对不同类型犯罪进行分析，分析不同犯罪类型的高发

区域和高发月份。因此这种分析方法有助于公安执法人员的警力部署和指挥，从而达到预防犯罪的目的。

参 考 文 献

- [1] Hitesh Kumar Reddy Toppi Reddy, Bhavna Saini, Ginika Mahajan. Crime Prediction & Monitoring Framework Based on Spatial Analysis. *Procedia Computer Science*, 2018(132): 696-705
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition//IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 770-778
- [3] Hu Tao, Zhu Xinyan, Duan Lian, Guo Wei. Urban crime prediction based on spatio-temporal Bayesian model. *PloS one*, 2018, 13(10): 1-18
- [4] 王增利, 刘学军, 陆娟, 吴伟, 张宏. 犯罪网络构建及其时空分析——以入室盗窃为例. *武汉大学学报(信息科学版)*, 2018, 43(05): 759-765
- [5] 段炼, 胡涛, 朱欣焰, 叶信岳, 王少华. 顾及时空语义的疑犯位置时空预测. *武汉大学学报(信息科学版)*, 2019, 44(05): 765-770
- [6] Martineau M, Beauregard E. Journey to Murder: Examining the Correlates of Criminal Mobility in Sexual Homicide. *Police Practice and Research*, 2016, 17(1): 68-83
- [7] Tayebi M A, Glasser U, Ester M, et al. Personalized Crime Location Prediction. *Eur. J. Appl. Math*, 2016, 27(3): 422-450
- [8] Nasridinov, A., Ihm, S. Y., & Park, Y. H.. A decision tree-based classification model for crime prediction//Information Technology Convergence, 2013: 531-538
- [9] Iqbal, R., Murad, M. A. A., Mustapha, A., Panahy, P. H. S., & Khanahmadliravi, N.. An experimental study of classification algorithms for crime prediction. *Indian Journal of Science and Technology*, 2013, 6(3): 4219-4225