

6-5-2020

Key Technologies of Precaution and Prediction of Abnormal Spatial-Temporal Trajectory: A Review of Recent Advances

Gongda Qiu

1. College of Command and Control Engineering, The Army Engineering University of PLA, Nanjing 210002, China;;

He Ming

1. College of Command and Control Engineering, The Army Engineering University of PLA, Nanjing 210002, China;; 2. Institute of network information, Academy of Systems Engineering, Academy of Military Sciences, Beijing 100071, China;;

Yang Jie

3. Science and Technology Information Office, Public Security Bureau of Jiangsu Province, Nanjing 210007, China;;

Yuting Cao

1. College of Command and Control Engineering, The Army Engineering University of PLA, Nanjing 210002, China;;

See next page for additional authors

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Key Technologies of Precaution and Prediction of Abnormal Spatial-Temporal Trajectory: A Review of Recent Advances

Abstract

Abstract: The ex-post disposition of a major incident, which is expected to transform into prediction and precaution of abnormal behavior, is increasingly unable to meet the urgent needs of the society. The rapid development and popularization of sensor network and positioning technology lay the foundation for mining spatial-temporal trajectory data. With the key objective of prediction and precaution of abnormal trajectory based on big data mining, *the future research directions and prospects on trajectory clustering and recognition are analyzed, discussed and elaborated in this paper. Temporal trajectory prediction applied in prediction and precaution of abnormal spatial-temporal trajectory is also presented*, providing a reference for further research on this field.

Keywords

spatial-temporal trajectory, abnormal behavior, clustering, prediction

Authors

Gongda Qiu, He Ming, Yang Jie, Yuting Cao, and Jihong Sun

Recommended Citation

Qiu Gongda, He Ming, Yang Jie, Cao Yuting, Sun Jihong. Key Technologies of Precaution and Prediction of Abnormal Spatial-Temporal Trajectory: A Review of Recent Advances[J]. Journal of System Simulation, 2017, 29(11): 2608-2617.

异常轨迹数据预警与预测关键技术综述

仇功达¹, 何明^{1,2}, 杨杰³, 曹玉婷¹, 孙继红^{1,4}(1.中国人民解放军陆军工程大学指挥控制工程学院, 南京 210002; 2. 军事科学院系统工程研究院网络信息研究所, 北京 100071;
3.江苏省公安厅科技信息化处, 南京 210007; 4. 南京市明基医院, 南京 210004)

摘要: 重大突发事件的事后处置已经愈加无法满足当前社会的迫切需求, 急于需要向事前异常行为的预警预测转型。传感器网络与定位技术的快速发展与普及, 为时空轨迹数据挖掘奠定了基础。围绕异常轨迹预警预测挖掘这一核心目的, 对异常轨迹聚类识别与轨迹预测的国内外研究现状和进展进行了理论梳理、剖析, 综述了相关算法在城市异常轨迹数据预警预测中的应用, 指出了所面临的挑战和进一步的发展方向, 为该领域的进一步研究提供参考。

关键词: 时空轨迹; 异常行为; 聚类; 预测

中图分类号: TP311

文献标识码: B

文章编号: 1004-731X (2017)11-2608-10

DOI: 10.16182/j.issn1004731x.joss.201711002

Key Technologies of Precaution and Prediction of Abnormal Spatial-Temporal Trajectory: A Review of Recent Advances

Qiu Gongda¹, He Ming^{1,2}, Yang Jie³, Cao Yuting¹, Sun Jihong⁴(1.College of Command and Control Engineering, The Army Engineering University of PLA,Nanjing 210002, China;
2.Institute of network information, Academy of Systems Engineering, Academy of Military Sciences, Beijing 100071, China;
3.Science and Technology Information Office, Public Security Bureau of Jiangsu Province, Nanjing 210007, China;
4.Nanjing BenQ hospital, Nanjing 210004, China)

Abstract: The ex-post disposition of a major incident, which is expected to transform into prediction and precaution of abnormal behavior, is increasingly unable to meet the urgent needs of the society. Therapid development and popularization of sensor network and positioning technology lay the foundation for mining spatial-temporal trajectory data. With the key objective of prediction and precaution of abnormal trajectory based on big data mining, the future research directions and prospects on trajectory clustering and recognition are analyzed, discussed and elaborated in this paper. Temporal trajectory prediction applied in prediction and precaution of abnormal spatial-temporal trajectory is also presented, providing a reference for further research on this field.

Keywords: spatial-temporal trajectory; abnormal behavior; clustering; prediction

引言

当前, 各类恐怖事件和突发事件频发, 通过事

件发生后产生的损失与危害触发的事后处置已经愈加无法满足当前社会稳定的迫切需求, 急于需要向事前异常行为的预警预测转型。与此同时, 随着视频监控网络、全球定位系统(GPS)、手持移动设备和射频识别(RFID)等设备的普遍应用。大数据在给人们带来大量信息的同时, 也面临着从中深度性地、关联性地发现有用知识的挑战。实现基于时空轨迹数据的异常行为发现与预警预测, 将有助于提



收稿日期: 2017-02-04 修回日期: 2017-07-04;
基金项目: 江苏省自然科学基金(BK20150721, BK20161469), 江苏省科技基础设施建设计划(BM2014391), 江苏省重点研发计划(BE2015728, BE2016904), 国家重点研发计划(2016YFC0800606), 中国博士后基金(2015M582786, 2016T91017);
作者简介: 仇功达 1(1992-), 男, 浙江余姚, 硕士, 研究方向为时空轨迹数据挖掘、公共安全。

<http://www.china-simulation.com>

• 2608 •

高社会治安防控能力。

(1) 基于轨迹聚类识别的异常行为预警技术

暴恐突发事件嫌疑对象在预谋、实施、藏匿等不同阶段所产生的行为多具有异常性,但轨迹又往往隐藏在众多繁杂的轨迹之中。当前研究重点围绕轨迹相似度描述、轨迹聚类算法,实现对异常行为轨迹的高效提取,为危险人员的精准识别提供支撑,做到有效的事前识别与预警。

(2) 基于历史或近似行为的轨迹预测技术

暴恐份子在预谋、实施、藏匿等不同阶段所产生的行为具有行为相似性、目的一致性。当前的轨迹预测重点围绕从历史轨迹^[1],或相似轨迹中建立

概率模型、机器学习模型实现预测,或通过轨迹聚类,从相似轨迹中检索出相关信息进行预测,研究者们改进并设计出了一系列的预测模型。但暴恐轨迹的异常性导致其难以从普通轨迹中获得较多的类似轨迹,有限的历史轨迹与近似轨迹集,制约了轨迹预测的准确性。

本文对时空轨迹数据的聚类识别预警、轨迹预测两个关键技术展开讨论,轨迹聚类不仅能识别模式识别异常,也能为预测圈定相应的学习数据,而轨迹预测能对轨迹进行补全,提高轨迹聚类的质量,两者之间存在密不可分的关系。框架图如图 1 所示。

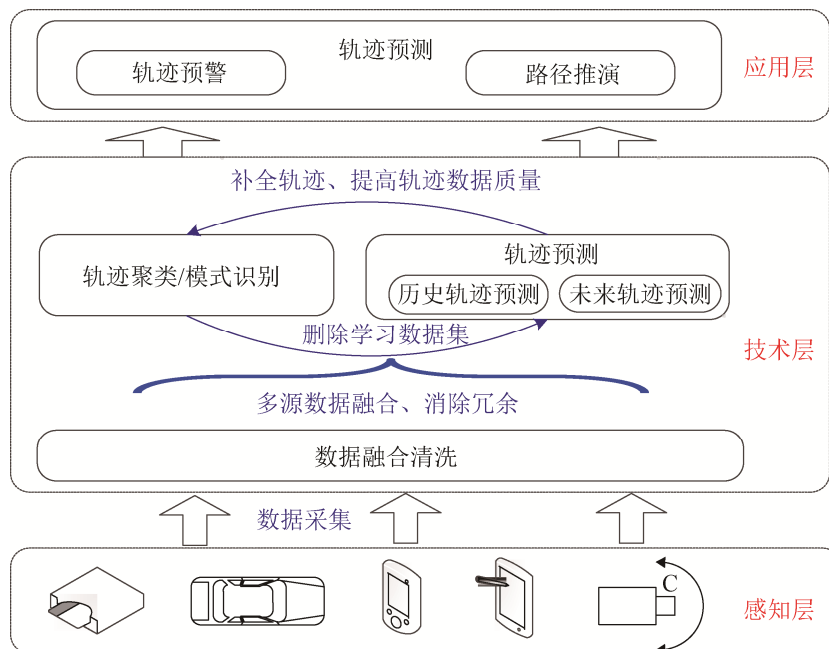


图 1 异常轨迹数据预警预测框架图

Fig.1 Frame of prediction and precaution of abnormal spatial-temporal trajectory

1 异常轨迹识别

轨迹异常检测^[2]能对隐藏在众多正常模式中的异常行为进行识别。Zhu 等^[3]将当前轨迹与历史轨迹对比,提出了一个基于轨迹孤立点分析的热门路径(TPRO)发现方法,该方法中与相应时段热门路径进行对比,将存在较大差异的轨迹视为异常;Chawla 等^[4]提出了一个两阶挖掘优化框架,通过检测与历史轨迹偏离程度定义异常,不仅关注异常路

径,并推断导致异常出现的路径;LEE 等^[5]提出了基于分段与分组密度聚类框架的检测算法,以“与大多数轨迹不存在最小要求相似距离”作为异常轨迹划分的判定标准。

以上可概括为在某一个距离描述下聚类,计算与热门轨迹的相异程度,识别离群轨迹。但在轨迹聚类上聚类精度有限,另外离群轨迹涵盖范围太宽泛,直接作为异常轨迹说服力不强,缺乏相关异常

信息挖掘补充。而其中的关键技术轨迹聚类识别研究的热点在于：相似度计算^[6-7]与聚类算法。

早期对轨迹的平面几何曲线进行聚类，以发现热点区域，Masciari^[8]提出了一种基于空间分割的聚类算法，根据移动目标的点位置、轨迹将搜索空间划分为合适的粒度。

但时间与位置的结合才有利于发现更加丰富的运动模型，因此希望尽可能的将各种轨迹信息包含起来，Mitsch 等^[9]总结到：时间维度中的取样时间瞬时值在很多研究只是用来发现线性模式，而其他时间属性，如循环时间模式或间隔在大多数情况下未处理的。Birant 和 Kut^[10]包括循环时间模式，Nanni 等^[11]考虑了时间间隔。

在这一过程中发现，许多的轨迹特征最后会被汇总到一个整体相似度上，导致掩盖了特征信息。Lee 等^[5]提出了一个分段与分组的轨迹聚类的框架。在他的研究中，提出了一个基于最小分割长度的正式轨迹分割以及子轨迹聚类算法。为了简化处理，yuan 等^[12]提出了根据实现设置的转向阈值实现拐角检测，来对轨迹进行分段。该类算法的优点在于可以对复杂高维轨迹有效检测出特征轨迹，然而，也容易受到分段结果的影响。

1.1 轨迹相似度测量

对不同属性的数据的相似度描述，以及不同需求下的聚类要求，产生了一系列的相似度描述方式，即距离计算公式。每个点的各维度或者说属性之间，是否有相关性，可以将距离分做闵可夫斯基距离与马氏距离。闵可夫斯基距离在指数参数的取值上，又可以获得欧氏距离、曼哈顿距离、切比雪夫距离，若将数据进行标准化，则获得的就是兰氏距离。当前轨迹相似度的计算大部分还是以欧式距离为基准，探索轨迹的结构，从结构上，结合欧式距离进行相似度描述。如豪斯多夫距离是计算了两条轨迹所有点间的欧式距离，取了两个单向距离中那个较大的值；DTW(Dynamic Time Warping)距离同样计算了两条轨迹所有点间的欧式距离，是通过

动态规整，进行了最小代价对齐。最小外包矩形，则是外轮廓的欧式距离。

(1)以局部相似情况作为相似度

豪斯多夫距离先求一个轨迹中的点到另一个轨迹中的所有点的最小距离，然后在所有最小距离中找到最大值。豪斯多夫距离即为轨迹间两个单向距离中那个较大的值。

Chen Jing-yang 等^[13]提出了一种基于改进的豪斯多夫距离的轨迹聚类算法。该算法首先豪斯多夫距离给出了种新的轨迹子段距离度量方法，以消除轨迹字段之间的公共偏差。然后利用特征点将轨迹划分成轨迹子段集，并计算轨迹子段之间的相似度，并由此实现聚类。

(2)以整体相似情况作为相似度

DTW 距离可较好地应对轨迹在时间维度上进行局部伸缩变化，解决了采样频率不稳定造成的轨迹数据时间尺度不统一的问题。另外，在计算动态时间封装距离时，在进行轨迹间的记录点映射时，要求轨迹具有连续性。因此总体上既具有一定的鲁棒性又对于离群值敏感。

Sankoff 等^[14]提出了一种针对时间扭曲、字符串编辑的序列比较方法，该方法中使用 DTW 距离来度量序列的相似性。但该方法的计算量过大。Little 等^[15]对轨迹数据的时空结构进行了分析，他们通过路径和速度曲线来表示轨迹，再用 DTW 方法来度量轨迹之间的距离。还有学者将轨迹引入极坐标空间，通过角度与长度来表示轨迹，再计算轨迹间的 DTW 距离。

(3)最快速、低精度且抗噪的距离计算

最小外包矩形可以用于简化时空轨迹。在处理轨迹数据的过程中，可以先将整条轨迹划分成若干子轨迹，再将每条子轨迹用最小的外包矩形来表示，这样每条轨迹就变成了一个最小外包矩形的序列，最后通过比较最小外包矩形序列即可度量时空轨迹间的相似性。郑宇等^[16]提到在大量轨迹处理时为显著提升处理效率时会以牺牲一定精度的代价来获得更快的处理速度。

(4)可处理数值化属性与语义属性的距离描述

当应用最长公共子序列^[17]描述轨迹间相似度时,如果是语义属性,则以是否相同为依据,若是数值型,以差值阈值作为依据,不要求整条轨迹相似,重点寻找不重叠的多个相似子区间,并将所有子区间的相似性汇总成轨迹间的相似性度量。针对轨迹数据的相似性度量,常用的是两条轨迹中的最长的相似子轨迹来衡量轨迹间的相似性,如下:

$$D_L(T_i T_j) = \begin{cases} 0; n = m = 0 \\ 1 + LCSS_{\sigma, \varepsilon}(\text{Head}(T_i), \text{Head}(T_j)); |a_i^x - b_j^x| \leq \sigma, \\ \text{and } |a_i^y - b_j^y| \leq \varepsilon \\ \max \max(LCSS_{\sigma, \varepsilon}(\text{Head}(T_i), T_j), \\ LCSS_{\sigma, \varepsilon}(T_i, \text{Head}(T_j))) \end{cases} \quad (1)$$

其中: σ 和 ε 分别表示 x 轴和 y 轴上的距离阈值。当两段轨迹的长度都为零时,轨迹间的最长公共子序列距离为零;当两者的起点在相似范围内时,轨迹间的最长公共子序列距离为除起点外剩下轨迹的最长公共子序列长度加 1;否则,在个轨迹与另一个轨迹的剩下部分中求一个最大的最长公共子序列长度作为最长公共子序列距离。

编辑距离(edit distance)^[18]指两个序列之间,由一个转成另一个所需的最少编辑操作次数。轨迹数据是由一个一个的染点组成的时间序列,因此可以很容易地将编辑距离的计算方法扩展到时空轨迹间的相似性度量上。两条轨迹 T_i 和 T_j 之间的编辑距离采用递归的方式进行计算,公式如下:

$$ED(T_i, T_j) = \begin{cases} n = m = 0 \\ m = n = 0 \\ ED(R(T_i), R(T_j)) m, n > 0 \text{ and } p_1^i = p_1^j \\ \min \begin{cases} ED(R(T_i), R(T_j)) + 1 \\ ED(R(T_i), T_j) + 1 \text{ otherwise} \\ ED(T_i, R(T_j)) + 1 \end{cases} \end{cases} \quad (2)$$

其中: m 和 n 分别代表时空轨迹 T_i 和 T_j 中含有的记录点的个数; p_1^i 和 p_1^j 分别表示轨迹 T_i 和 T_j 之上的第一个点间; $R(T)$ 则表示轨迹中除去第一个点后

剩余的轨迹。当其中一条轨迹的记录点数目为零时,它们之间的编辑距离就是另一条轨迹中记录点的个数;如果存在相同的起点,则它们之间的编辑距离就是剩下的轨迹的编辑距离;如果没有相同的起点,则编辑距离增加,通过递归的方式求取最小值作为两条轨迹之间的编辑距离。

1.2 轨迹聚类算法

(1)密度聚类算法

Palma 等^[19]用基于密度的聚类算法发现了轨迹中的兴趣点。在他的研究中,通过应用基于改进的 Eps 线性近邻距离函数的 DBSCAN 算法对目标的移动与静止进行了删除。LEE 等^[5]提出了一个分段与分组的轨迹聚类的框架。在他的研究中,先是提出了一种轨迹分割方法,将轨迹表示为众多子轨迹的集合,最后对子轨迹段进行聚类,其中所采用的聚类算法正是 DBSCAN 算法。Nanni 与 Pedreschi^[11]提出了将 OPTICS 算法应用到轨迹数据聚类中的 T-OPTICS 算法。OPTICS 算法与 DBSCAN 算法具有相似的理论,轨迹聚类结果的基础上,时间主要是勾勒出时间维度的内在语义,以便提高轨迹聚类的质量。

(2)其他密度聚类算法

CFSFDP^[20](clustering by fast search and find of density peaks)则可视作是 DBSCAN 的改良版,通过密度与可达距离分离密度类峰值点,实现快速聚类。

DENCLUE(density based clustering)算法^[21]则是基于一个全局密度分布函数的聚类算法,聚类可以通过确定全局密度函数的局部最大点来梯度可达。设定统一的路径阈值 a , 以及峰值阈值 E , 对于阈值需精心选择,这些参数会显著影响聚类情况。并且当面对多密度聚类时,稀疏与密集区域的阈值的选择会发生显著冲突。

这些算法在应对多密度情况下数据分布时均会显得局促。为绝对多密度数据分布问题,主要的算法有 Chameleon^[22]、共享近邻 SNN 算法^[23]、多

阶段等密度线算法等^[24]。

(3) 针对大数据集改良的快速聚类

陈黎飞等^[25]针对在大型数据集上计算效率欠佳问题,提出了基于层次划分的最佳聚类数确定方法。构建了一条关于不同层次划分的聚类质量曲线,曲线极值点所对应的划分用于估计最佳的聚类数目。CF Tsai, ZC Chen, CW Tsai 等^[26]结合多重搜索遗传算法(MSGA)与 K-means 算法提出了一种新的多重搜索遗传 K 均值算法(MSGKA),可以将给定数据的全局最优分割发现到指定数量的簇中。王莉等人^[27]针对 Lingras 提出的粗 K 均值聚类算法中存在的易受随机初始聚类中心和离群点的影响而出现一致性和无法收敛的问题,对此提出一种改进的粗 K 均值算法。

2 路径推演与轨迹预测技术

路径推演与轨迹预测从最典型的概率图模型出发,根据状态转移预测时是否可逆,即有向无向性将概率图模型分做两类,如图 2 所示。无向性代表有马尔科夫随机场,仅与各维度变量邻近状态有关。而在贝叶斯体系下,状态转换概率具有有向性。在贝叶斯体系下又分为静态贝叶斯与动态贝叶斯。静态指基础模型、转换概率并不发生改变,主要有混合模型、自回归。而动态贝叶斯网络中主要包括:隐马尔科夫模型、马尔科夫链模型、卡尔曼滤波模型,转换概率或模型结构会根据输入的动态变化而进行变化。

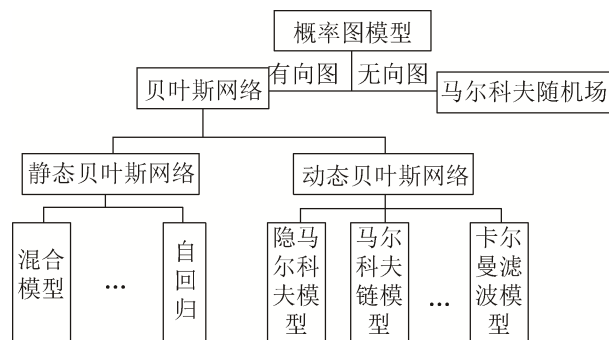


图 2 概率预测模型分类图

Fig. 2 Classification of probability prediction model

从预测内容上可分为两类,一是关注位置状态转移的语义预测,二是基于具体坐标的经纬度预测。图 3 所示为该分类下目前大多数的预测文章所采取的模型,在此基础上展开进行改良,或结合实际中众多限制改良。从路网限制上可以分为路网下预测与非路网限制预测,路网下轨迹存在突变转向,在轨迹连续性上的突变会使得一般模型难以准确预测,但是优点在于结合路网信息,可以对轨迹进行修正。从数据源上分,则可分为基于历史轨迹与基于相似轨迹的轨迹预测。

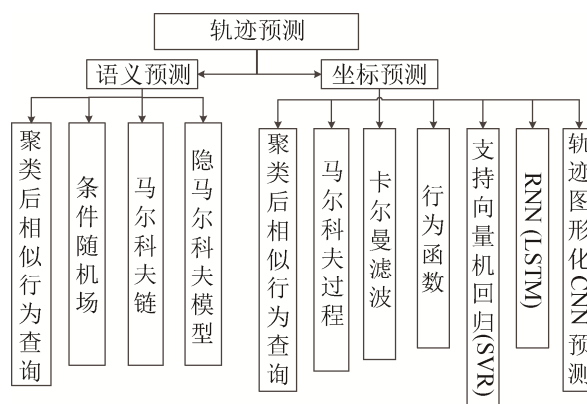


图 3 常见轨迹预测方法

Fig. 3 Common trajectory prediction methods

2.1 基于马尔科夫性质的轨迹预测

在轨迹预测中具有众多基于马尔科夫性质的预测模型:马尔科夫链(针对离散时间、离散状态),马尔科夫过程,隐马尔科夫模型(状态不直接可见)。鉴于马尔科夫性质:未来状态仅与当前状态有关,而轨迹预测中并非全是仅基于当前状态的轨迹预测,因此衍生出了保留 n 步历史状态的 n 阶马尔科夫链。

Gambs 等^[28]把马尔科夫链拓展为融合 n 步历史状态的 MMC 模型(Mobility Markov Chain),基于历史轨迹采样点对移动对象的下一个位置进行预测,但在路网限制下,历史轨迹的效用发挥有限。

Marmasse 和 Schmand^[29]在研究中也使用了隐马尔科夫模型来预测用户的下一个位置,并将其与贝叶斯分类器进行对比,发现统计模型的预测结果

具有更好的效果。通过深入分析得出如下结论,在训练数据量较少的情况下,统计模型得出的效果将稳定且优于其他方法。

2.2 基于其他概率模型的轨迹预测

朴素贝叶斯推理预测模型,主要由训练阶段和轨迹预测阶段两部分组成。训练阶段,将历史轨迹或横向类似行为轨迹作为训练数据集,对预测模型进行学习,构造出模型;预测阶段,基于训练得到的模型,对查询轨迹进行分析和预测。可以给出形式化的公式,如下:

$$P(l_n \in c_j | T^q) = \frac{P(l_n \in c_j | T^q) P(l_n \in c_j)}{\sum_{1 \leq k \leq \text{NumC}}^{c_k} P(l_n \in c_j | T^q) P(l_n \in c_j)} \quad (3)$$

这里的先验概率 $P(l_n \in c_j)$ 可以用数据库中包含 l_n 的轨迹数目除以数据库中所有轨迹数目。当历史数据库中不包含某位置 l_n 的轨迹时,求得的先验概率将变为零。即无法实现对未知或未曾走过的路段的轨迹预测。

Jeung 等^[30]提出了一个混合预测模型,对 Apriori 算法进行改进并且通过对移动对象运动模式的挖掘,把移动对象的运动行为抽象成复杂的数学公式,并结合这个行为函数进行未来轨迹的预测。但是如何在公式中融入人的意图,且基于路网的轨迹并非连续可导函数,因为路口选择,轨迹具有突变性。函数构建存在困难。

Roland Assam 和 Thomas Seidl^[31]提出了一个采用条件随机场(CRF)的预测模型。基于前后信息聚类 and 社交关系来预测下一个位置。首先基于用户相似性和时间段的相似性的评价准则计算任意两个重要位置(即由相邻的驻点组成的地区)的相似度。运用谱聚类构造图模型,将相似度比较大的重要位置相连构成一个聚类。最后基于条件随机场模型使用上述构建的位置集群和用户移动性行为预测下一个位置。

2.3 基于聚类查询的轨迹预测

轨迹聚类识别技术,将有助于为目标轨迹删选相关合适的数据集作为统计学习、神经网络学习等的基础数据集。在处理后数据集的基础上实现对异常轨迹的下一步轨迹预测。在横向上选择在限定属性下的相似轨迹作为补充,纵向上对目标历史轨迹进行聚类删选,寻找当前条件下的历史近似行为,作为数据集的补充。

Morzy 等^[32]提出了 Traj-PrefixSpan 算法,对移动对象轨迹的频繁模式进行挖掘构建概率模型以预测未来轨迹。但该方法没有考虑轨迹时间等相关特性。Ozer 等^[33]增加了通话日志记录数据,他们首先对通话数据进行聚类处理以增大数据粒度、减少数据量,然后他们提出基于时空数据的频繁模式挖掘的预测算法,该算法预测位置的变化,而不再是预测个别对象的下一个位置,在执行效率和精度上有着优势,且较为适合路网下的预测。

Zhengyu 等^[34]提出了一种 TBHG 模型,用于模拟多个用户的历史位置。TBHG 模型由具有基于树的层次结构 H,以及树中每一层次的图模型 G,这两个结构的整合。树结构代表着重要位置之间的父子级别关系,同时图模型表达的是节点之间的对等关系。基于这个结构,运用连接分析算法(HITS)推断用户的经验旅行路线,以及对各轨迹点的兴趣度。以出行时间轨迹序列为输入,进行下一个轨迹位置预测。

2.4 基于神经网络的轨迹预测

斯坦福 AlexandreAlahi 等^[35]借鉴递归神经网络(RNN)模型在序列预测上的成功,提出了一个 social-LSTM 模型来学习一般人体运动轨迹并预测他们在未来时刻的路径选择问题。不仅考虑了单个轨迹在时序上的演变过程,还关注了当几条轨迹同时发生时,周围轨迹对某一轨迹的影响。AlexandreAlahi 等的工作对当前许多的时序预测问题具有较好的启迪作用。

JianmingLv 等^[36]则是利用卷积神经网络(CNN)

提出了 T-CONV 模型, 通过构建不同精度的网格化的多尺度空间, 采用卷积神经网络提取特征, 实现对出租车轨迹进行精确的坐标预测。

2.5 各种针对性改良下的轨迹预测

针对干扰消除问题, 陈凌等^[37]提出了 CRPM 算法通过对连续路径模式挖掘来避免其他轨迹对目标轨迹预测的干扰, 在此基础上进行未来轨迹的预测。但是没有在预测中考虑其他属性, 例如时间属性和运输方式。Pelekis 等^[38]考虑轨迹数据库中固有的不确定性对轨迹数据挖掘的影响, 基于直觉模糊基对 C 均值算法进行改进提出了用于处理不确定性轨迹的聚类方法 CenTR-I-FCM, 该方法有效的解决了对不确定性轨迹的聚类, 但是考虑的对比参数仅是空间上的相似性度量。

针对精度低、信息利用率低、开销大等问题, Lei P R 等^[39]提出了 STT 模型以获取移动对象的运动行为, 还可以减少用于存储轨迹模式的大小, 并基于 STT 模型提出了 STT 轨迹预测算法, 对未来轨迹的位置进行估计, 不仅可以利用空间信息来进行预测, 而且还可以利用时间信息来估计目标位置, 进一步提高了轨迹预测的准确性。

针对轨迹数据采样频率低的情况, Chiang M F 等^[40]提出了基于轨迹点可达性预测模型的受时间限制的移动图, 用以预测采样频率低并且要求预测未来很长时间的轨迹。Doug Cox 等^[41]提出了一种改良马尔科夫链蒙特卡洛方法算法针对低采样频率、低精度运动传感器网络来进行轨迹推理。并计划对马尔科夫链返回的一系列结果构建一个信心指标, 并用一个期望最大化算法来完善初步环境模型, 以在线更新的方式获得更准确的传输矩阵与时间分布模型。

3 未来研究展望

3.1 现有问题的分析与展望

时空轨迹数据具有十分重要的应用价值, 已成为世界范围的研究热点。从目前国内外研究现状分

析发现, 国外已开发出用于实践验证的移动对象轨迹预警预测系统, 但是在大规模实际应用还存在一定的距离。与国外发达国家相比, 国内对时空轨迹数据方面的研究成果无论从数量还是深度, 都尚有差距。异常轨迹的预警方面, 尚停留在热点模式的识别, 对异常的挖掘能力不足, 另外关于时空轨迹数据的预测问题, 国内研究也更多停留在马尔科夫链的改进, 对新的预测模型的探索不足。现有的这些方法不能完全满足行业应用的需求, 目前的研究存在以下几点不足。

1) 异常轨迹识别问题。

①当前轨迹聚类主要关注于频繁模式的挖掘, 实现路线规划提示。在聚类中重点完成大类即频繁行为的识别, 而对小类轨迹的识别缺少进一步的提升, 如何尽可能精确的识别出更多的轨迹类, 提升聚类算法在应对复杂数据分布时的识别精度与识别速度。

②距离描述缺少突破。轨迹相似性的度量较多基于欧式空间下的距离度量, 更多从轨迹结构进行探索, 构造相似度描述方式。没有从维度相关性、高维信息融入、整体相似度与局部特征矛盾等角度开展。缺少对属性的进一步分析。包括属性的重要性、属性相关性等等, 信息属性增加难。

③异常的定义不够充分。较多研究均将离群值作为了异常, 缺乏对所有离群值的进一步异常信息挖掘。可以将静态一次聚类识别转换为多种代表距离下的动态轨迹聚类, 通过在不同信息侧重下的聚类结果获得多层次的理解, 生成更加丰富的信息。例如从时间维度, 当轨迹阶段性增长后, 聚类结果的动态变化中进行识别分析。

2) 轨迹预测问题

①轨迹预测模型缺少突破。当前轨迹预测尤其国内非常多是基于马尔科夫链或其他改良模型的改良, 或是通过历史轨迹相似轨迹的查询补充。对基于目标主体喜好行为习惯、环境信息的学习模型研究较少。基于概率模型的预测在少量轨迹下的预测较为占优, 深度学习模型又缺少数据及其他信息

的支撑。

②缺少精度上的突破。也正是由于精度预测的问题,更多实用的预测是基于语义状态转移预测。对基于具体坐标的位置预测研究不少,但成效不显著。

③路网的限制。在无路网下的轨迹,其行为与目的具有连贯性,在轨迹一开始就会表现出意图,预测上精度更高。而基于路网的轨迹,速度方向等状态存在突变,正是由于路网的限制,其轨迹意图与行为并不连续。导致众多预测模型在路网未修正前精度较差。

④信息利用少。当前众多时序状态的预测仅是基于坐标信息,从历史坐标点中去解读下一步的行为意图。忽略了移动目标主体的实时决策属性。可以尝试结合更多的信息,如路网拥堵系数、转向频率习惯,还可以辅以相似路径对历史轨迹进行模糊化,以消除局部异常导致的误差。

3.2 其他的可能方向

除上述所述不足外,基于时空轨迹数据的预警预测研究还可以在以下几个方面进行尝试:

(1)高维轨迹聚类验证难。高维轨迹无法直接观测,对于一定精度以上的聚类存在验证难等问题。在实际应用过程中,人对实验结果的监督理解,至关重要,将对决策提供支持与辅助。可进一步探索将高维数据向低维的映射过程,投影到人能理解的二三维空间,亦可从多个维度进行切片投影,以反映某些属性下的数据分布。数据维度变化对聚类挖掘具有重要意义。

(2)轨迹聚类缺乏尺度探索。轨迹预警作为辅助决策技术,不仅包括核心聚类异常识别,包括聚类结果展示。而不同尺度上,轨迹的各个特性比重并不相同,面对时间跨度短、轨迹自身较短。分布地域广的轨迹聚焦数据,更多呈现的是轨迹的重心分布引起的差别,而不是轨迹形状特性带来的差距。不同时间跨度,轨迹起始时刻引起的差异也将不同。等等当前众多轨迹聚类算法并未考虑尺度信

息,直接进行距离,这将导致不同的数据集将带来完全不同的聚类效果。

(3)路网限制下的转向预测。路网下的轨迹预测可以尝试通过对路口信息、历史路口信息等的学习实现转向预测,当转向与路网结合,实际上已经完成了轨迹。路网下路口的状态与直行行驶的状态重要性并不等价。

(4)轨迹相互影响模型下的预测问题。众多的轨迹预测模型将训练集轨迹一条条地单独输入模型中,或增加环境信息进行训练,斯坦福的 Alahi A 在其 social-LSTM 模型中不仅考虑了轨迹自身性质及所处环境信息,还考虑了周围轨迹对其产生的影响。启示了我们不应仅关注对轨迹单独的训练学习,也应关注轨迹间的相互影响,建立更加全面的模型。

4 结论

时空轨迹数据研究正处于发展阶段,是计算机学科与交通、环境、经济、能源、社会安全和城市规划等多个学科的交汇。本文重点关注在异常轨迹预警预测方面应用,系统性地总结了预警预测关键技术的国内外研究现状和发展情况;对目前的研究进行了分析和问题总结并对未来研究方向进行了展望,希望为基于时空轨迹数据的异常行为预警预测研究提供新思路。

参考文献:

- [1] Hariharan R, Toyama K. Project Lachesis: parsing and modeling location histories[M]. Adelphi, MD: Springer Berlin Heidelberg, 2004: 106-124.
- [2] 毛嘉莉, 金澈清, 章志刚, 等. 轨迹大数据异常检测: 研究进展及系统框架[J]. 软件学报, 2017, 28(1):17-34. Mao Jiali, Jin Cheqing, Zhang zhigang, et al. Anomaly Detection for Trajectory Big Data: Advancements and Framework[J]. Journal of Software, 2017, 28(1):17-34.
- [3] Zhu J, Jiang W, Liu A, et al. Time-Dependent Popular Routes Based Trajectory Outlier Detection[C]// International Conference on Web Information Systems Engineering. Springer International Publishing, 2015:16-30.

- [4] Chawla S, Zheng Y, Hu J. Inferring the Root Cause in Road Traffic Anomalies[C]//IEEE, International Conference on Data Mining. IEEE, 2012:141-150.
- [5] Lee J G, Han J, Whang K Y. Trajectory clustering: a partition-and-group framework[C]//ACM SIGMOD International Conference on Management of Data. ACM, 2007:593-604.
- [6] Lee S L, Chun S J, Kim D H, et al. Similarity Search for Multidimensional Data Sequences[C]// International Conference on Data Engineering, 2000. Proceedings. IEEE, 2000:599-608.
- [7] Toohey K, Duckham M. Trajectory similarity measures[J]. Sigspatial Special(S1946-7729), 2015, 7(1): 43-50.
- [8] Masciari E. A framework for trajectory clustering[C]// International conference on GeoSensor Networks. Springer Berlin Heidelberg, 2009: 102-111.
- [9] Mitsch S, Müller A, Retschitzegger W, et al. A Survey on Clustering Techniques for Situation Awareness[M]//Web Technologies and Applications. Springer Berlin Heidelberg, 2013: 815-826.
- [10] Birant D, Kut A. ST-DBSCAN: An algorithm for clustering spatial-temporal data[J]. Data & Knowledge Engineering(S0169-023X), 2007, 60(1): 208-221.
- [11] Nanni M, Pedreschi D. Time-focused clustering of trajectories of moving objects[J]. Journal of Intelligent Information Systems(S0925-9902), 2006, 27(3): 267-289.
- [12] Yuan G, Xia S, Zhang L, et al. An efficient trajectory-clustering algorithm based on an index tree[J]. Transactions of the Institute of Measurement &Control (S0142-3312), 2012, 34(7):850-861.
- [13] Chen Jing-yang, Song Jia-tao, Liu Xu-liang, et al. Trajectory Clustering Algorithm Based on Improved HausdorffDistance[J]. computer engineering (S1000-3428), 2012, 38(17):157-161.
- [14] Sankoff D, Kruskal J B. Time warps, string edits, and macromolecules : the theory and practice of sequence comparison[M]. Addison-Wesley Pub. Co. Advanced Book Program, 1983.
- [15] Little J J, Gu Z. Video retrieval by spatial and temporal structure of trajectories[C]//Photonics West 2001 - Electronic Imaging. International Society for Optics and Photonics, 2001:545-552.
- [16] Zheng Y. Trajectory Data Mining: An Overview[J]. Acm Transactions on Intelligent Systems &Technology (S2157-6904), 2015, 6(3):1-41.
- [17] 魏龙翔, 何小海, 滕奇志, 等. 结合 Hausdorff 距离和最长公共子序列的轨迹分类[J]. 电子与信息学报, 2013, 35(4):784-790.
- Wei Long-xiang, He Xiao-hai, Teng Qi-zhi, et al. Trajectory Classification Based on Hausdorff Distance and Longest Common SubSequence[J]. Journal of Electronics & Information Technology, 2013, 35(4): 784-790.
- [18] Chen L, Ng R T. On the marriage of Lp-norms and edit distance[C]. The Vldb Journal, 2004: 792-803.
- [19] Palma A T, Bogorny V, Kuijpers B, et al. A clustering-based approach for discovering interesting places in trajectories[C]//Proceedings of the 2008 ACM symposium on Applied computing. ACM, 2008: 863-868.
- [20] Rodriguez A, Laio A. Machine learning. Clustering by fast search and find of density peaks[J]. Science (S0036-8075), 2014, 344(6191):1492-1496.
- [21] Hinneburg A, Gabriel H H. DENCLUE 2.0: Fast Clustering Based on Kernel Density Estimation[C]// International Conference on Intelligent Data Analysis. Springer-Verlag, 2007:70-80.
- [22] Shang F, Jiao L C, Shi J, et al. Fast density-weighted low-rank approximation spectral clustering[J]. Data Mining & Knowledge Discovery(S1384-5810), 2011, 23(2):345-378.
- [23] Pan Zhang-ming, Chen Yi-li. Research on Shared Nearest Neighbor Clustering for Large Dataset[J]. Journal of Chinese Computer Systems(S1000-1220), 2014, 35(1): 50-54.
- [24] Wu Yu-hong. General Overview on Clustering Algorithms[J]. Computer Science(S1002-137X), 2015, 42(S1).
- [25] 陈黎飞, 姜青山, 王声瑞. 基于层次划分的最佳聚类数确定方法[J]. 软件学报, 2008, 19(1): 62-72.
- Chen Lifei, Jiang Qingshan, Wang Shengrui. A Hierarchical Method for Determining the Number of Clusters [J]. Journal of Software, 2008, 19(1): 62-72.
- [26] Tsai C F, Chen Z C, Tsai C W. MSGKA: an efficient clustering algorithm for large databases[C]//IEEE International Conference on Systems, Man and Cybernetics. IEEE, 2003, 6: 5.
- [27] 王莉, 周献中, 沈捷. 一种改进的粗 K 均值聚类算法[J]. 控制与决策, 2012, 27(11):1711-1714.
- Wang Li, Zhou XianZhong, ShenJie. An improved rough k-means clustering algorithm [J]. Control and Decision, 2012, 27(11): 1711-1714.
- [28] Killijian M O. Next place prediction using mobility Markov chains[C]//Proceedings of EuroSys 2012

-

- [71] Pan J, Manocha D. GPU-based parallel collision detection for fast motion planning[J]. Journal of Robotics Research (S0278-3646), 2012, 31(2):187-200.
- [72] Gunther J, Popov S, Seidel H P, et al. Realtime ray tracing on GPU with BVH-based packet traversal[C]// Proceedings of IEEE Symposium on Interactive Ray Tracing, 2007: 113-118.
- [73] Aila T, Laine S. Understanding the efficiency of ray traversal on GPUs[C]//Proceedings of High Performance Graphics, 2009: 145-149.
- [74] Zhang X Y, Kim Y J. Scalable collision detection using p-Partition fronts on many-core processors[J] IEEE Transactions on Visualization and Computer Graphics (S1077-2626), 2014, 20(3): 447-456.
- [75] Tang M, Tong R F, Wang Z D, et al. Fast and exact continuous collision detection with Bernstein sign classification[J]. ACM Transactions on Graphics, 2014, 33(6): 186.
- [76] Dynamic Benchmarks[DB/OL]. <http://gamma.cs.unc.edu/DYNAMICB/>.
- [77] FracturingModels[DB/OL]. <http://sglab.kaist.ac.kr/models/>.