基于生存分析的GPS轨迹缺失规律挖掘

郑剑炜 顾晶晶 庄 毅
(南京航空航天大学计算机科学与技术学院 南京 211100)

摘 要 近年来,智能交通系统(Intelligent Transportation Systems,ITS)已成为提高交通系统性能和增强出行安全性的有效方式。但随着系统数据量的增加,数据缺失问题日益严重,其中由于车载GPS信号丢失导致的轨迹数据缺失是主要的研究问题之一。引起GPS轨迹缺失的原因的多样性造成数据补全工作困难,且至今很少有关于轨迹缺失规律的研究。针对GPS信号丢失原因多样化的问题,基于大量真实数据,首次将生存分析应用于数据缺失领域,提出了基于生存分析的GPS轨迹缺失规律挖掘模型(Survival Analysis-Missing Trajectory Pattern Mining,SA-MTPM)。首先通过生存函数描述信号丢失时长与丢失原因的关系,然后利用Cox回归模型分析信号丢失的关键因素。使用上海市强生出租车公司一个月内13666辆车的数据进行实验,结果表明GPS轨迹缺失存在一定规律,据此可以方便地对信号丢失事件进行识别分类,为进一步对大数据进行研究提供了参考。

关键词 轨迹缺失,信号丢失,生存分析,规律挖掘
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.05.031

Pattern Mining of Missing GPS Trajectory Based on Survival Analysis

ZHENG Jian-wei GU Jing-jing ZHUANG Yi
(College of Computer Science and Technology,Nanjing University of Aeronautics and Astronautics,Nanjing 211100,China)

Abstract In recent years,intelligent transportation systems(ITS) has been an effective way to improve the traffic performance of transportation system and enhance the safety of travels. However,with the increase of data size in intelligent transportation system,the problem of data loss becomes increasingly serious. The trajectory data missing caused by vehicle-mounted GPS signal loss is one of the main research subjects. The reasons of GPS data missing are various,and they make the data completion difficult. However,there are few studies on the pattern of missing GPS trajectories. In this paper,based on large amounts of real data on diversification of GPS signal loss,the survival analysis was first applied into data missing field,and a survival analysis-missing trajectory pattern mining(SA-MTPM) model was proposed. The relationship between the length of signal loss and the regression causes of loss was described in the survival function,and the Cox model was used to analyze the key factors of signal loss. This paper performed experiments based on the GPS data of 13666 vehicles in Shanghai Qiangsheng Taxi Company for a month. The experimental results show that these signal loss events can be classified,which provides a further study for big data.

Keywords Track missing,Signal loss,Survival analysis,Pattern mining

1 引言

近年来,传感器技术、通讯技术和计算能力得到迅速发展,现实社会中人们的行为被数字化后,形成了海量的数据。这些数据集的规模及其多维性使得学者有机会探索许多社会行为规律,并基于这些规律进行有效的预测。基于此产生了城市智能交通系统(ITS),其具有非常大的研究潜力。现代城市中大量车辆(例如出租车)配备了GPS设备,使得城市智能交通系统可以利用采集到的大规模数据来挖掘并分析城市的动态与车辆轨迹间隐藏的规律,从而实现路线规划软件、交通流预测[1]、异常报警系统[2]和智能交通信息平台等多样化的应用创新。

基于车辆GPS数据的研究涵盖许多方面,例如轨迹补全、轨迹异常检测等。然而,当车辆处于复杂环境时,设备自身的不稳定性将造成GPS数据的缺失。通过观察本文使用的数据集发现,缺失数据往往存在于城市热点区,例如火车站、机场等。城市热点区数据的丢失将导致很难开展一些ITS应用研究,例如城市功能块划分、智能车辆异常检测以及数字地图绘制等。

现有的处理缺失数据的方法主要利用已知数据对缺失数据进行估计,从而实现补全。Rubin[3]提出了一种缺失数据的类型学,将丢失数据分为3类:完全随机丢失、随机缺失和非

到稿日期:2017-03-14 返修日期:2017-06-05 本文受国家自然科学基金面上项目(61572253),航空基金项目(2016ZC52030)资助。
郑剑炜(1993—),男,硕士生,主要研究方向为模式识别;顾晶晶(1983—),女,副教授,主要研究方向为模式识别、无线传感器网络,E-mail:gujingjing@nuaa.edu.cn(通信作者);庄 毅(1956—),教授,博士生导师,主要研究方向为网络与分布式计算、信息安全、可信计算。

随机丢失[4]；Pommeret 等[5]利用参数化处理来检验数据丢失机制，但此方法的实用性较差；孙婕等[6]给出了 3 种数据丢失机制的检验识别方法；Shan 等[7]提出了基于多模型方法的长期车辆运动预测和跟踪算法；Li 等[8]利用多个点的时间和空间的相关性，提出了改进的基于主成分分析与内核概率主成分分析的补全算法；Asif 等[9]提出了监督的学习方法，如 k 均值聚类、主成分分析和自组织图，这些方法挖掘网络层面中每个个体链接的时空性能趋势；Mitrovic 等[10]提出在缺少历史和相邻数据的情况下可以构造大型和多样网络的低维表示方法，该方法主要使用低维模型重建路段的数据概况并估算缺失值。

由于车载 GPS 设备所处环境复杂，具有缺失原因多样和缺失样本量大的问题，直接利用时-空相关性进行数据补全时效果不佳，因此需要对数据缺失样本进行分类处理，然而目前很少有文献研究挖掘 GPS 轨迹缺失的规律。

生存分析是一个统计分支，用于调查事件与时间之间的规律，但目前没有文献应用生存分析来挖掘车辆轨迹丢失规律。在生存分析的相关工作中，Schonfelder[11]利用生存分析来探索基于旅行日记的旅行行为模式；May 等[12]利用生存分析来处理 GPS 调查中的缺失数据；王冠男等[13]提出了基于生存分析的照片轨迹时空规律挖掘，大大简化了置信区间求解以及假设检验的过程；环梅等[14]利用生存分析对非机动车闯红灯行为进行了研究，并对此类行为的规律进行了分析；孙剑等[15]提出了城市快速路瓶颈交通流失效生存分析方法，针对快速路常发性瓶颈失效的随机特征建立了生存分析模型。

本文首次利用生存分析的方法，基于数据丢失的时间长度与丢失事件的关系来挖掘 GPS 轨迹缺失的规律，并应用一种来自生存分析领域的技术（即 Kaplan-Meier 估计）来对其进行评估；同时，运用 Cox 回归模型来分析 GPS 轨迹基本参数对信号丢失事件的影响。

本文第 2 节介绍了生存分析方法，并说明了生存分析在本文背景下的应用；第 3 节介绍了如何运用生存分析对 GPS 轨迹缺失规律进行挖掘；第 4 节在真实数据集上进行实验，验证了本文所提方法的有效性；最后总结全文并展望未来。

2 背景与问题描述

生存分析主要用于医学、生物学和工程学领域。Cheung 等[16]利用生存分析方法得出肝切除术与开放肝切除术的单中心经验；Hammouche[17]进行了非典型脑膜瘤的长期生存分析，得到了放射治疗与其他因素之间的关系；Musci 等[18]对青年首次吸烟的时间进行生存分析，评估了多因素对吸烟时间的干预影响。

生存分析考虑了从一个起始事件到兴趣事件的时间。这里的事件是指在临床研究中出现的一些疾病或在质量控制中出现的装置故障。一种典型的生存分析方法是 Kaplan-Meier，其在给定时间段内估计一些事件不发生的概率（即感兴趣的对象存活），允许存在删失数据。删失数据指对象的生存时间存在删失，一般存在于兴趣事件未发生或对象丢失时。例如，在癌症治疗过程中，患者从开始治疗到死亡的时间为该对象的生存时间，兴趣事件为患者死亡。若患者在治疗期间突然失联或者在规定时间间隔内未死亡，则产生数据删失。在生存分析的过程中经常会遇到完全数据和不完全数据。本文中生存时间指的是车辆轨迹缺失的时间，即一辆车从丢失 GPS 信号到恢复信号所经历的时间。因此，完全数据代表在测量时间内车辆轨迹从丢失变为恢复，而不完全数据代表未检测到轨迹恢复。

本文中 GPS 数据缺失的原因主要有：1）路过隧道等卫星信号无覆盖区域，如图 1 所示，车辆经过外滩隧道时完全丢失信号，GPS 呈现为隧道入口 S 到出口 D 的直线；2）路过飞机场、火车站等信号干扰区；3）设备老化导致接收信号不灵敏；4）外力造成设备关闭。

图 1 上海“外滩隧道”GPS 轨迹缺失实例

Fig. 1 Example of track missing in Shanghai Waitan tunnel

3 基于生存分析的 GPS 轨迹缺失规律的分析与挖掘

生存分析能够通过分析生存时间来挖掘不同因素对事件的影响规律，其效率与准确性高，适用于大数据，因此本文首次将生存分析方法运用于挖掘数据缺失规律，提出了基于生存分析的 GPS 轨迹缺失规律挖掘模型 SA-MTPM。首先通过生存函数描述信号丢失时间长度与丢失原因的关系，然后利用 Cox 回归模型分析信号丢失的关键因素。

3.1 GPS 轨迹数字化分析

原始车辆信息数据集无序，且含有大量的无效数据，需对数据集进行预处理，使最终获得的轨迹符合生存分析要求。首先，根据车辆 ID 对数据集进行分类，当遇到格式有误或者信息缺失的记录时，删除记录或者适当补全记录；其次，按时间顺序将同一 ID 车辆的 GPS 信息串联成连续 GPS 点，进而形成车辆轨迹；最后，在图 2 给出的上海市的电子地图中，将城市地图分成相等大小的网格单元，对所有网格单元进行编号，将每一条 GPS 轨迹与网格相匹配，记录经过的网格序号，从而形成连续的符号序列，轨迹效果如图 3 所示。

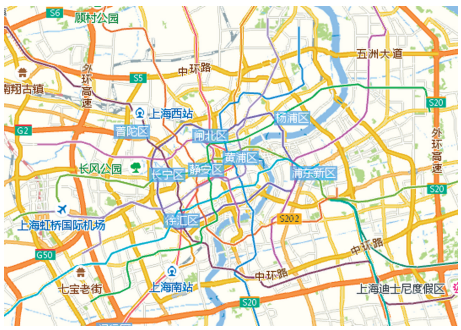

图 2 上海市区路网图

Fig. 2 Shanghai road network

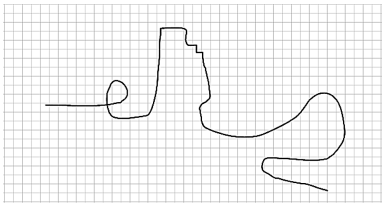

图 3　一辆出租车在 2 小时内的轨迹

Fig. 3　Trajectory of a taxi within 2 hours

基于文献[19]，GPS 信号丢失的情况大多可在数据集中体现，其特点为丢失点与其临近点的时间或距离差超过一定阈值。由于 GPS 设备的精度差别较大、测量时间间隔各异，没有理论最佳值，同时市区中 GPS 信号丢失的情况复杂，不易判断，因此将两个临近测量点的时间间隔作为阈值。本文所使用数据集的理论测量时间间隔为 5 s，由于存在通讯延迟等问题，根据经验将阈值定为 6 s，即当两个临近测量点的时间间隔超过阈值 6 s 时，判断其为信号丢失的情况。当检测到某轨迹存在信号丢失情况时，记录丢失路段的起点与终点的网格序号。

3.2　GPS 轨迹缺失的生存分析

3.2.1　GPS 轨迹缺失生存模型的构建

在完全数据与不完全数据并存的数据集中，构造 GPS 轨迹缺失生存模型的关键在于建立相应的生存函数。令 T 为对象生存时间的一个随机变量，即兴趣事件发生的时间。生存函数如式(1)所示[20]：

$$S(t)=P(T>t) \tag{1}$$

其中，$P(T>t)$ 表示指定事件发生时间晚于某个时间 t 的概率。

令 r_i 表示在 t_i 时刻对象的总数量，e_i 表示发生兴趣事件的对象数量，c_i 表示在每一次测量时间间隔内退出的对象数量。规定在 t_0 时刻，$e_0=c_0=0$。在事件发生的每个时间点，对象的总数量因兴趣事件发生以及对象的退出而减少，即 $r_{i+1}=r_i-e_i-c_i$。K-M 估计通过计算两个连续事件之间的条件概率来适应不同的样本量。假设在两个事件之间退出研究的对象生存，直到下一个事件发生时才被删除，则在 t_i 时刻，该对象存活的条件概率如式(2)所示：

$$p_i=P(T>t_i\mid T>t_{i-1})=\frac{r_{i-1}-e_i}{r_{i-1}} \tag{2}$$

设 $T_1,T_2,\cdots,T_n$ 为本文中的测量时间间隔，其为非负且相互独立的随机变量，它们具有同一生存函数；基于不同时刻的条件概率为 p_i。则某一时刻 t_k 的生存概率如式(3)所示：

$$S(t_k)=P(T>t_k)=\prod_{i=1}^{k}\frac{r_{i-1}-e_i}{r_{i-1}} \tag{3}$$

在生存分析中，事件发展的过程是对象从一个状态到另一个指定状态的转移过程，可定义危险率函数来描述任意时刻状态转移的速度。危险率函数如式(4)所示：

$$h(t)=\frac{\lim_{\Delta t\to\infty}(P(t\leqslant T<t+\Delta t)\mid T\geqslant t)}{\Delta t} \tag{4}$$

生存函数 $S(t)$ 也称作可靠度函数，其相对应的死亡函数 $F(t)$ 如式(5)所示：

$$F(t)=P(t\leqslant T)=1-S(t) \tag{5}$$

利用上述公式对 GPS 轨迹缺失生存函数进行估计的步骤如下：

Step1　根据观察值大小对总量为 r_0 的样本进行排序，由小到大依次为 $y_0,y_1,\cdots,y_n$，若存在观察值相等的完全数据与不完全数据，则将完全数据放置在前；

Step2　列出不同时间间隔 $T_1,T_2,\cdots,T_n$ 下 t_i 时刻 r_i，e_i 和 c_i 的值；

Step3　计算不同时间间隔下在 t_i 时刻指定对象存活的条件概率 p_i 以及危险率 $h(t)$；

Step4　根据式(4)计算某一时刻 t_k 的生存概率 $S(t_k)$ 以及累积死亡率 $F(t)$；

Step5　利用计算得到的不同时刻的 $S(t_k)$ 和 $h(t)$ 形成累计生存率曲线与累计危险值曲线，通过比较曲线分布来对轨迹缺失类型进行区分。

上述步骤可以对不同的数据丢失类型进行分类，但是无法量化其他因素对丢失事件产生的影响。由于在复杂环境下存在多种协变量，因此引入 Cox 比例风险回归模型来评估协变量对事件的影响程度。

3.2.2　Cox 比例风险回归模型

通过生存函数计算得到的结果只取决于对象的生存时间，而忽略了其他的因素。Cox 比例危险模型可以应用于不完全数据，可以分析其他因素对对象的生存时间产生的影响，属于半参数方法。Cox 回归模型引入了协变量，其对生存分析中的危险率函数与基准风险函数的比值(即 $\frac{h(t)}{h_0(t)}$)产生影响并反映在数值变化中。Cox 比例风险回归模型如式(6)所示：

$$\begin{aligned}h(t,X)&=h_0(t)\cdot f(X,\beta)\\&=h_0(t)\exp(\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p)\end{aligned} \tag{6}$$

其中，$h(t)$ 为 t 时刻对象的危险率函数；$h_0(t)$ 为基准风险函数；$\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p$ 为预后指数；$\beta_i(i=1,2,\cdots,p)$ 为函数的回归系数。回归系数可为正也可为负，当回归系数为正数时，表示该参数会对生存增加风险值，从而使生存时间下降，即该参数是一个危险因子，生存时间随回归系数的增大而递减；当回归系数为负数时，表示该参数会减少生存的风险值，从而使生存时间延长，即该参数是保护因子。基准风险函数 $h_0(t)$ 是指所有协变量都为标准 0 状态时的风险函数，其与时间相关且一般为未知。假定所有对象的危险函数成比例，则比例系数如式(7)所示：

$$h(t)/h_0(t)=\exp(\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p) \tag{7}$$

其中，基准风险函数 $h_0(t)$ 未定义，是半参数模型。比例系数反映了不同变量对信号丢失的影响，其中的参数可以通过大量的实际样本得到的观察值来确定。

4　实验

4.1　数据来源

本文采用的数据集为真实数据集，来源于 2015 年 4 月 1 日至 2015 年 4 月 30 日共 30 天的上海市强生出租车 GPS 轨迹数据，每辆出租车的数据信息细节如表 1 所列，其中载客信息有 0 和 1 两个指标，0 表示空闲，1 表示载客。

表 1 车辆数据信息细节

Table 1 Taxi dataset details

车辆 ID	载客	时间	经度	纬度	速度/(km/h)	方向角/°
30043	1	2015/04/01 13:21:52	121.452047	31.118147	25	196
30043	1	2015/04/01 13:21:57	121.451847	31.117515	33	194

实验数据整体情况如表 2 所列。对于丢失的 GPS 轨迹，本文基于原始数据集与高德地图开发平台，对每条轨迹的丢失类型进行可视化判别，丢失类型分为 4 种：区域丢失、信号干扰、设备老化和外力关闭。

表 2 实验数据

Table 2 Experimental data

数据集	2015 年 4 月上海强生出租车的 GPS 信息
总出租车数/辆	13666
有效天数/天	30
采集间隔/s	5
GPS 信号丢失百分比/%	3

4.2 Kaplan-Meier 参数估计

由于实验数据中存在删失数据，因此分别对不考虑删失数据和考虑删失数据两个数据类型进行生存分析建模，图 4 为在两种情况下的累积生存曲线对比图。在 50% 的累计生存率(即 50% 的出租车恢复信号)的情况下，考虑删失数据和不考虑删失数据的累积生存曲线对应的 GPS 信号丢失时间分别为 30 s 和 60 s，两种情况下的累积生存曲线存在明显差异。因此，本文的生存分析模型考虑删失数据。

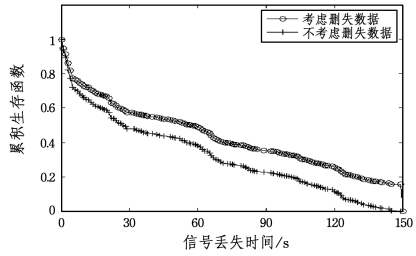

图 4 有、无删失数据情况下的生存曲线对比图

Fig. 4 Comparison of survival curves with or without censored data

确定考虑删失数据后，对 4 种信号丢失类型的数据分别进行生存分析，分析结果如图 5 所示。

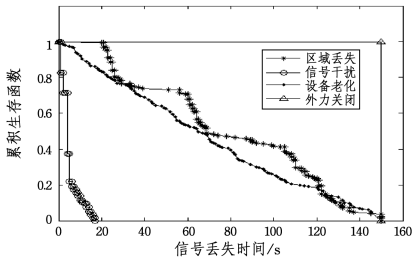

图 5 4 种信号丢失类型的生存曲线图

Fig. 5 Survival curves for four signal loss types

由图 5 可以看出，4 种信号丢失类型具有明显差异。当车辆 GPS 轨迹处于区域丢失时，累积生存曲线在 25 s，60 s 和 120 s 处下滑快速，这反映出在实际环境中车辆经过 3 条不同隧道而丢失 GPS 信号；当 GPS 轨迹由于信号干扰而丢失时，累积生存曲线整体下滑迅速，车辆均于 20 s 内恢复 GPS 信号，在不同丢失时间下危险率变化无明显差异；当车辆由于设备老化丢失 GPS 信号时，累积生存曲线下滑均匀，危险率基本保持不变，这反映出由于设备老化造成的信号丢失并无规律；当 GPS 信号由于外力关闭而丢失时，累积生存函数在 150 s时由 1 变到 0，说明此类数据均为删失数据，这些删失数据由于超过测量时长而截尾。

4.3 Cox 半参数估计

该实验研究 GPS 信号丢失类型与设备使用年龄对信号丢失时长产生的影响。应用 Cox 回归模型进行单因素分析。

将 3 种 GPS 信号丢失类型代入 Cox 模型进行分析，结果如表 3 所列。

表 3 GPS 信号丢失类型的 Cox 单因素回归分析

Table 3 Cox proportional hazard regression for GPS signal loss types

因素	回归系数 β	标准误差 SE	P	RR	95%CI
区域丢失	−0.011	0.108	0.031	0.989	0.800～1.223
信号干扰	−3.927	0.228	0.000	0.020	0.013～0.031
外力关闭	1.873	0.143	0.000	6.509	4.914～8.622

从实验结果可以看出，上述 3 种 GPS 信号丢失类型(区域丢失、信号干扰和外力关闭)均具有统计学意义，是影响信号丢失时间的关键因素，且 3 种类型的差异明显，可以通过分析结果明确分辨信号的丢失类型。

将设备使用年龄代入 Cox 模型进行分析，结果如表 4 所列。

表 4 设备使用年龄的 Cox 单因素回归分析

Table 4 Cox proportional hazard regression for device age

因素	回归系数 β	标准误差 SE	P	RR	95%CI
$age=1$	0.070	0.178	0.694	1.072	0.757～1.519
$age=2$	−0.151	0.158	0.337	0.859	0.631～1.171
$age=3$	−0.085	0.156	0.584	0.918	0.676～1.247
$age=4$	−0.034	0.156	0.828	0.966	0.711～1.313
$age=5$	−0.130	0.156	0.402	0.878	0.648～1.190

从实验结果可以看出，$P_{age}>0.05$，$age=1,2,\cdots,5$，因此设备使用年龄不具有统计学意义。为了更清楚地显示设备使用年龄对生存时间的影响，对不同情况下的累积生存函数进行分析，结果如图 6 所示。显然，现有使用年龄在 1～5 年内的 GPS 设备的累积危险函数曲线变化并无明显差异，因此设备使用年龄不是出租车 GPS 信号丢失的关键因素。

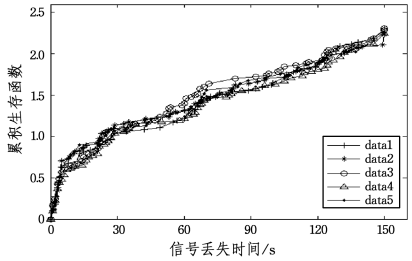

图 6 不同设备使用年龄的累积危险函数曲线

Fig. 6 Cumulative hazard function curves for different device ages

结束语　本文针对车辆 GPS 设备信号丢失的问题，在大量 GPS 真实数据的基础上识别并分离出丢失的行车路段，构建了 GPS 信号丢失事件的生存分析模型，并利用 Cox 回归模型对影响 GPS 丢失时间的关键因素进行了分析，通过实验挖掘出了一些车辆 GPS 轨迹缺失的规律。结果表明，与传统统计学方法相比，本文的分析方法能够考虑失效样本与删失数据，分析结果更加全面。通过生存分析发现，GPS 轨迹缺失存在一定规律，不同信号丢失类型（区域丢失、信号干扰、设备老化和外力关闭）的生存曲线存在明显差异，可以利用此规律方便地对现有丢失事件进行分类；当设备使用年龄在 5 年内时，不同设备并无明显差异。未来将在已有信号丢失规律的基础上对轨迹进行分类，以进一步完善城市电子路网和 GPS 热点区域。本文工作为拓展 ITS 应用，例如智能电子数字地图、城市模块识别规划软件、无人车应用等提供了便利。

参 考 文 献

[1] LV Y, DUAN Y, KANG W, et al. Traffic Flow Prediction With Big Data: A Deep Learning Approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2): 865-873.

[2] FANASWALA M, KRISHNAMURTHY V. Detection of Anomalous Trajectory Patterns in Target Tracking via Stochastic Context-Free Grammars and Reciprocal Process Models[J]. IEEE Journal of Selected Topics in Signal Processing, 2013, 7(1): 76-90.

[3] RUBIN D B. Inference and missing data[J]. Biometrika, 1976, 63(3): 581-592.

[4] SCHAFER J L, GRAHAM J W. Missing data: our view of the state of the art[J]. Psychological Methods, 2002, 7(2): 147-177.

[5] OMMERET D. Testing the mechanism of missing data[EB/OL]. [2017-03-23]. https://hal. archives-ouvertes. fr/hal-00669339.

[6] SUN J, JIN Y J, DAI M F. Discussion on the Test Method of Data Deletion Mechanism [J]. Mathematics in Practice and Theory, 2013, 43(12): 166-173. (in Chinese)
孙婕，金勇进，戴明锋. 关于数据缺失机制的检验方法探讨[J]. 数学的实践与认识，2013，43(12)：166-173.

[7] SHAN M, WORRALL S, NEBOT E. Probabilistic Long-Term Vehicle Motion Prediction and Tracking in Large Environments [J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(2): 539-552.

[8] LI L, LI Y, LI Z. Efficient missing data imputing for traffic flow by considering temporal and spatial dependence[J]. Transportation Research Part C Emerging Technologies, 2013, 34(9): 108-120.

[9] ASIF M T, DAUWELS J, GOH C Y, et al. Spatiotemporal Patterns in Large-Scale Traffic Speed Prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(2): 794-804.

[10] ASIF M T, MITROVIC N, GARG L, et al. Low-dimensional models for missing data imputation in road networks [C]// International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013: 3527-3531.

[11] SCHÖNFELDER S, AXHAUSEN K. Analysing the rhythms of travel using survival analysis[C]// Transport Research Board (TRB) 2001 Annual Meeting. 2000.

[12] MAY M, KÖRNER C, HECKER D, et al. Handling missing values in GPS surveys using survival analysis: a GPS case study of outdoor advertising[C]// International Workshop on Data Mining & Audience Intelligence for Advertising. ACM, 2009.

[13] WANG G N. Spatial-Temporal Data Mining Based on GPS Trajectory and Geo-Tagged Photo Trajectory [D]. Changsha: Central South University, 2013. (in Chinese)
王冠男. 基于 GPS 轨迹和照片轨迹的时空数据挖掘[D]. 长沙：中南大学，2013.

[14] HUAN M, YANG X B, JIA B. Red-Light Running Behavior of Non-Motor Vehicles Based on Survival Analysis [J]. Transactions of Beijing Institute of Technology, 2013, 33(8): 815-819. (in Chinese)
环梅，杨小宝，贾斌. 基于生存分析方法的非机动车闯红灯行为研究[J]. 北京理工大学学报，2013，33(8)：815-819.

[15] SUN J, ZHANG J. Survival Analyses of Traffic Flow Breakdown at Urban Expressway Bottlenecks [J]. Journal of Tongji University (Natural Science), 2013, 41(4): 530-535. (in Chinese)
孙剑，张娟. 城市快速路瓶颈交通流失效生存分析[J]. 同济大学学报自然科学版，2013，41(4)：530-535.

[16] CHEUNG T T, POON R T, YUEN W K, et al. Long-term survival analysis of pure laparoscopic versus open hepatectomy for hepatocellular carcinoma in patients with cirrhosis: a single-center experience[J]. Annals of Surgery, 2013, 257(3): 506.

[17] HAMMOUCHE S, CLARK S, WONG A H, et al. Long-term survival analysis of atypical meningiomas: survival rates, prognostic factors, operative and radiotherapy treatment [J]. Acta Neurochirurgica, 2014, 156(8): 1475-1481.

[18] MUSCI R J, FAIRMAN B, MASYN K E, et al. Polygenic Score× Intervention Moderation: an Application of Discrete-Time Survival Analysis to Model the Timing of First Marijuana Use Among Urban Youth[J]. Prevention Science, 2015, 27(1): 1-9.

[19] XIE K, NING X, WANG X, et al. Recover Corrupted Data in Sensor Networks: a Matrix Completion Solution [J]. IEEE Transactions on Mobile Computing, 2017, PP(99): 1.

[20] 彭非，王伟. 生存分析[M]. 北京：中国人民大学出版社，2004.