

数据-模型协同驱动的货车物流模式分类和验证*

许钧奇 王宁 张义龙

(同济大学汽车学院, 上海 201804)

【欢迎引用】许钧奇, 王宁, 张义龙. 数据-模型协同驱动的货车物流模式分类和验证[J]. 汽车文摘, 2025(3): 42-50.

【Cite this paper】XU J Q, WANG N, ZHANG Y L. Classification and Verification of Truck Logistics Modes Driven by Data-Model Collaboration[J]. Automotive Digest (Chinese), 2025(3): 42-50.

【摘要】我国公路货运市场存在运力分散化的问题,需要网络货运平台根据货运经营者出行偏好针对性分配物流任务,优化车、货供需匹配。为此,提出一种数据-模型协同驱动的货车物流模式分类方法。首先,基于货车轨迹数据构建回转半径、活动熵、日均出行地点、日均行驶时间、日均行驶距离5个特征参数,经主成分分析(PCA)降维及K均值聚类将货车物流模式划分为长途往返型、短途定点型、短途多点型。然后,引入图论中网络基序识别技术,通过货车OD数据生成有向物流网络,利用DotMotif算法进行基序识别,并选用 p 值检验基序显著性。最后,通过深入分析网络基序与货车出行链典型拓扑结构之间的联系,解释了不同物流模式货车的物流网络中显著性基序的差异,验证了货车物流模式分类结果的准确性。

关键词: 轨迹数据; 物流模式; K-means 聚类; 网络基序; 出行链

中图分类号: U495 文献标志码: A DOI: 10.19822/j.cnki.1671-6329.20240227

Classification and Verification of Truck Logistics Modes Driven by Data-Model Collaboration

Xu Junqi, Wang Ning, Zhang Yilong

(School of Automotive Studies, Tongji University, Shanghai 201804)

【Abstract】In China's road freight market, there exists the issue of decentralized transportation capacity, requiring online freight platforms to allocate logistics tasks according to travel preferences of freight operators and optimize the matching of vehicles and goods. Thus, a data-model collaborative-driven classification method for truck logistics patterns is proposed. Firstly, based on the truck trajectory data, 5 characteristic parameters including turning radius, activity entropy, average daily travel locations, average daily travel time, and average daily travel distance are constructed. After dimensionality reduction by Principal Component Analysis (PCA) and clustering by K-Means Cluster, trucks are categorized into 3 types of logistics patterns: long-distance round-trip, short-distance fixed-point, and short-distance multi-point. Secondly, the network motif identification technology from graph theory is introduced, generating a directed logistics network from truck OD data, using the DotMotif Algorithm for motifs recognition and selecting p -value to test the significance of motifs. Finally, by deeply analyzing the connection between network motifs and the typical topological structures of truck travel chains, the differences in significant motifs within the logistics networks of different logistics modes of trucks are explained, verifying the accuracy of the the truck logistics pattern classification results.

Key words: Trajectory data, Logistics patterns, K-means cluster, Network motif, Travel chain

0 引言

公路货运是我国货运的主力军,其份额占据我国整体货运量的75%,市场规模超5万亿^[1]。然而,目前国内公路货运市场中个体货运经营者占比超90%,运

力呈现明显的分散化和碎片化,导致货运效率低下。尽管各类网络货运平台一定程度上缓解了运力资源闲置问题,但以个体货运经营者为主的公路货运特征依旧显著。个体货运经营者的出行决策受自身出行偏好影响,形成了不同的物流模式,这也成为影响运

*基金项目:中央高校基本科研业务费专项资金资助(2023-4-YB-04)。

力分配和利用的关键^[2]。要有效地利用网络货运平台实现运力的集中管理和规划,必须识别出个体货运经营者偏好的物流模式,为其分配更适合的物流订单,从而提升车货匹配的精准度。

随着车载卫星定位装置普及,产生了海量的GPS轨迹数据,为研究货车物流模式提供了新思路^[3]。货车轨迹数据中包含车辆的经纬度、时间戳、速度等信息,通过挖掘其中蕴含的轨迹时空特征,可分析货车出行行为,支撑物流模式识别与分类。该方法已在人群活动模式研究中得到应用,例如,关庆锋等^[4]通过手机GPS数据构建惯性矩、职住距离、出行速度等特征参数,通过聚类分析研究了深圳市不同经济水平人群的活动模式。而在货车物流模式研究领域,此方法则应用较少,相关研究所构建的用于聚类分析的特征参数也较为简单,不能全面反映货车时空活动特征。例如, Ma等^[5]通过分析货车起止点(Origin-Destination, OD)数据,分析货车的平均出行时长、距离和停车时间分布等特征,将货运个体出行模式分为访问出行、本地出行及环形出行。Duan等^[6]从GPS数据中提取平均停留点数、平均停留时间、平均出行时间等特征,采用改进K-means算法得到5种出行模式。然而,直接利用轨迹数据中的时空特征分析货车物流模式忽略了实际物流网络结构,在揭示货车驾驶员个性化的物流模式偏好上存在局限性。

此外,建立出行链模型来分析货车物流模式的研究也较为常见。货运出行指货车从一个位置出发,中间无装卸货物直接到达另外一个位置^[7-8]。基于货运出行的概念,货运出行链被定义为货车进行一系列出行而形成的具有时间序列的链结^[8-9]。Ma等^[10]使用货车3天短期轨迹数据构建单次货运活动出行链,并使用非层次聚类算法将货车出行链的不同划分为4类,但由于数据集太小,未对货车物流模式进行进一步分析。针对此问题, You等^[11]使用时间跨度为1年的轨迹数据构建清洁货车出行链,挖掘出清洁货车在物流模式上的偏好。Gan等^[12]结合兴趣点(Point of Interest, POI)类型从轨迹数据中识别出各类型经停点,并以此绘制出主要出行链,将货运活动划分为走廊模式、环形模式和枢纽辐射模式。上述方法虽然可以很好地分析各类货车间的物流模式差异,但出行链的构建依赖于经停点的精确识别分类,当货车出行范围广、时间长、数量大时,构建出行链的繁琐程度大幅上升。

通过梳理文献发现,在基于轨迹时空特征数据的货车物流模式分类方法中,特征参数的构建较为简单,

而出行链模型视角下的货车物流模式分类方法依赖于精确的停留点提取,流程繁琐复杂。为此,本文提出一种数据-模型协同驱动的货车物流模式分类方法。首先,构建能够准确量化货车轨迹离散程度、空间范围、物流频率等特征的参数。然后,将特征参数经主成分分析(Principal Component Analysis, PCA)提取主成分用于K-means聚类输入,实现对货车物流模式的分类。最后,本文引入图论中的网络基序识别技术,对由货车OD数据构建的有向物流网络进行基序识别和显著性检验,并将网络基序结构与货车出行链典型拓扑结构相关联,通过分析不同聚类内部货车在网络基序结构上的共性,对不同物流模式货车中的显著基序的差异进行解释,验证货车物流模式分类结果的准确性。

1 数据处理

在货车行驶过程中,受GPS设备信号不稳定和地理环境等因素影响,采集的数据会出现缺失、重复、漂移等各种问题。因此,需对异常数据进行处理,并从源数据中提取OD数据,为后续聚类分析及基于网络基序的物流模式验证提供数据支持。

1.1 源数据分析

本文数据来源于某公司的货车车联网平台,主要分布于江西省。数据集包括自2023年3月1日至3月31日93辆货车的3 025 547条轨迹数据,采样间隔为10 s,源数据具体格式如表1所示。每条记录包括车架号、采集时间、经纬度、车速、发动机状态等多个字段。

表1 轨迹数据字段描述

字段	数据名称	数据单位
user_id	车架号	
time	采集时间	s
longitude	经度	(°)
latitude	纬度	(°)
velocity	车速	km/h
engine_speed	发动机转速	r/min
engine_state	发动机状态	
mileage	续航里程	km

对随机抽取的某辆货车A的轨迹数据进行可视化,结果如图1所示,可以从货车A的行驶轨迹中观察到其物流范围集中在东经112.0°~115.5°、北纬26°~29°。

1.2 缺失及异常值处理

源数据采集频率为10 s/次,若相邻数据采集时间相差过大,则存在整条数据缺失,此外,某些数据的部

分字段也存在缺失值。为了确保数据在时间序列上的完整性,本文采用移动平均法对源数据中的缺失值进行填补。同时,对于速度、经纬度等字段出现离群值的异常点也应先删除再采用移动平均法进行填补。

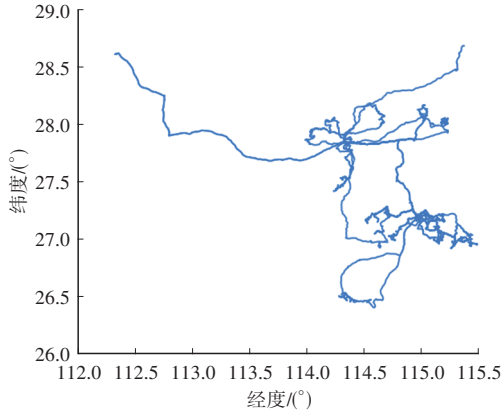


图1 单辆货车的原始轨迹可视化结果

1.3 漂移数据处理

数据漂移是指GPS测量位置与实际位置存在较大偏差的现象,理论上货车不可能在相邻两次采样中出现超过一定阈值的位置变化。实际数据中部分相邻数据的经纬度相差过大,造成虚假的行程,影响货运活动的识别。因此,需要对轨迹数据是否出现漂移进行检查,剔除漂移数据并采用移动平均法进行补充。

1.4 OD数据识别

OD数据是从源数据中提炼出的简化形式,记录了车辆每段行程的起始点、终止点的经纬度和时间。本文将发动机状态的变化作为识别行程起始和结束的关键指标,但在货车实际行驶过程中,可能由于红绿灯、堵车等原因致使发动机状态发生变化,而这些变化点并非期望识别到的OD点,因此,需结合货车时间、空间等数据进行分析,具体流程如图2所示。

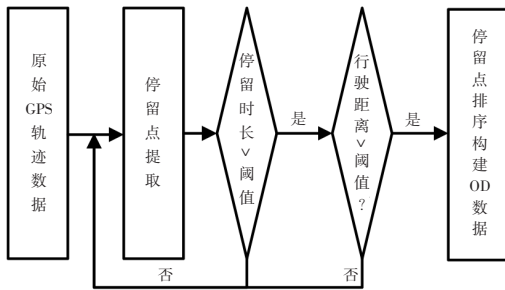


图2 轨迹数据OD识别流程

(1)停留点提取^[13]。预处理后的货车轨迹数据表示为 $\{l_1, l_2, \dots, l_m\}$,其中每条数据 $l_i = (x_i, y_i, T_i, \dots)$ ($i \in \{1, 2, \dots, m\}$)。对轨迹数据中以0或1标志的发动机状态参数进行一阶差分,可以初步得到货车的起讫点,即潜在停留点集合 $\{s_1, s_2, \dots, s_n\}$,其中

$s_i = (x_i, y_i, T_i^a, T_i^l)$ ($i \in \{1, 2, \dots, n\}$), x_i 和 y_i 表示潜在停留点的经纬度, T_i^a 和 T_i^l 表示停留点的到达时间和离开时间。

(2)停留时长判断。对每个停留点 s_i 计算离开时间和到达时间的差值,获取货车在每个停留点的停留时长 dur_i ,删去 dur_i 小于给定时间阈值的点。

(3)行驶距离判断。对停留时长满足要求的停留点排序,计算相邻停留点之间的行程距离 d_i ,进而筛选出 d_i 大于给定距离阈值的OD节点,得到行程 $trv_i = \{s_i, s_{i+1}\}$ 。

对每辆货车的源数据进行如上步骤,最终获取实验所需的OD数据,形式为^[8]:

$$OD = s_1 \xrightarrow{trv_1} s_2 \xrightarrow{trv_2} \dots \xrightarrow{trv_{N-1}} s_n \quad (1)$$

OD数据样例如表2所示,记录了起始点、终止点的经纬度和时间。

表2 OD数据样例

开始时间	O点经度/(°)	O点纬度/(°)	终止时间	D点经度/(°)	D点纬度/(°)
2023/3/1 8:56	116.974 7	32.597 8	2023/3/1 9:56	117.379 7	32.538 4
2023/3/1 9:57	117.379 8	32.538 4	2023/3/1 9:57	117.379 8	32.538 4
2023/3/1 10:56	117.379 8	32.538 4	2023/3/1 12:34	117.995 1	31.987 1
2023/3/1 13:22	117.995 1	31.987 0	2023/3/1 13:30	117.991 7	31.986 2

2 研究方法

2.1 基于轨迹特征的物流模式分类

2.1.1 特征参数构建

为了准确量化货车与物流模式相关的轨迹特征,本文构建五个特征参数,分别为回转半径、活动熵、日均出行地点^[14]、日均行驶距离和日均行驶时间,具体如表3所示。

表3 轨迹特征参数汇总

参数名称	符号	单位
回转半径	R_g	(°)
日均出行地点	A	
活动熵	H	
日均行驶里程	t	km
日均行驶时间	L	s

(1)回转半径:回转半径 R_g 表示货车活动范围半径, R_g 越大说明货车活动范围越大:

$$R_g = \sqrt{\frac{1}{M} \sum_{i=1}^M ((x_i - \bar{x})^2 + (y_i - \bar{y})^2)} \quad (2)$$

式中: M 为每辆货车轨迹数据的总采集条数, (x_i, y_i) 表示第*i*次采集时的货车纬度和经度坐标位置信息, (\bar{x}, \bar{y}) 表示整个货车轨迹网络的质心坐标。

(2)日均出行地点:日均出行地点*A*表示货车平均每日所停留的活动位置数量,*A*越大说明货车服务地点和客户需求点更多:

$$A = \frac{|\text{set}(s_1, s_2, \dots, s_n)|}{30} \quad (3)$$

式中: $\text{set}(s_1, s_2, \dots, s_n)$ 表示停留点的集合,仅包含不重复的停留点。

(3)活动熵:活动熵*H*表示货车的物流节点的离散程度:

$$p_i = \frac{\sum dur_j}{\sum_{j=1}^N \sum dur_j} \quad (4)$$

$$H = -\sum_{i=1}^N p_i \log(p_i) \quad (5)$$

式中: N 为总的停留点数量; $\sum dur_j$ 为货车在同一停留点*s_i*的停留时长总和; p_i 为货车在停留点*s_i*的停留时长占总体时长的比例, $\sum p_i = 1$ 。

(4)日均行驶时间:日均行驶时间*t*表示货车平均每日货运出行耗时:

$$t = \frac{T_i - \sum_{j=1}^N \sum dur_j}{30} \quad (6)$$

式中: T_i 为轨迹数据的总采集时间。

(5)日均行驶距离:日均行驶距离*L*表示货车平均每日货运出行距离:

$$L = \frac{\sum_{i=1}^n d_i}{30} \quad (7)$$

式中: d_i 为每段行驶段的行驶距离。

2.1.2 轨迹特征聚类

由于各参数之间存在相关关系,不必使用所有特征参数作为聚类输入,并且过多的特征参数会增加聚类分析难度,因此,本文采用PCA方法对数据进行降维。

为从货车轨迹中提炼出货车的物流模式,本文以PCA提取的主成分作为变量进行K-means聚类分析,K-means算法基于数据到中心点距离最小化的思想对数据进行无监督地聚类。本文利用SC轮廓系数衡量聚类效果,SC轮廓系数描述聚类后各个类别的轮廓清晰度,本质上表征的是聚类的内聚度和分

离度。SC轮廓系数值越小,说明该样本点所在的簇结构与其最近簇结构之间的距离越远,表示此时的聚类结果越好。

2.2 基于网络基序的物流模式验证

本文引入图论中的网络基序识别技术^[15],首先基于OD数据构建物流网络,然后应用基序识别算法识别网络中的基序。为检验这些基序的出现是否显著,本文为每个物流网络模拟生成500个分布相似的随机网络作为对照组,通过统计随机网络中相应基序的预期出现频数进行显著性检验。进一步地,结合货运出行链,深入分析不同物流模式货车的物流网络中显著性基序的差异,可验证物流模式分类结果的准确性。

图论是研究图的性质和特征的学科,在图论中,图被定义为由节点和边组成的对象。而复杂网络将复杂系统中的实体抽象成节点,将实体之间的关系抽象成边,因此,可以利用图的结构和性质来理解复杂网络中的问题。在将复杂网络表示为图时,可以认为网络基序是复杂网络中出现频率较高、具有特定拓扑结构的子图,这些基序反映了复杂网络中局部的组织结构和交互方式,对于识别网络中的重要结构、揭示网络中的重要信息和机制具有重要意义。

2.2.1 物流网络构建

利用货车OD数据构建物流网络的具体步骤如图3所示。首先,对OD数据进行栅格化,定义栅格尺寸为200 m,从而将距离相近的OD点归类到一个栅格。然后,为简化网络结构的同时保留货运活动信息,对相同起讫点的分支进行识别,并通过判断分支流量是否小于阈值来决定是否合并,当合并分支时需验证轨迹连通性,若合并分支后造成轨迹不连通,则取消合并、保留分支。最后,确定货车物流网络的节点和边的集合,通过向空的有向图中添加节点和边构造有向网络。

2.2.2 随机网络生成

为了识别基序在网络中的显著性,需将基序在目标网络中的出现频率与随机网络中的预期出现频率进行对照,因此,要求生成的随机网络必须与目标网络具有相同的节点数,且每个节点入度和出度相同。本文借助NetworkX库中Configuration Model来生成指定节点出度和入度的随机网络,Configuration Model是一种从给定度序列生成随机网络的方法,其将节点度数序列中的每个节点拆分为对应数量的半边,然后将这些半边随机配对,构成完整的边。本文通过Configuration Model为每辆货车物流网络生成分布相

似的500个随机网络。

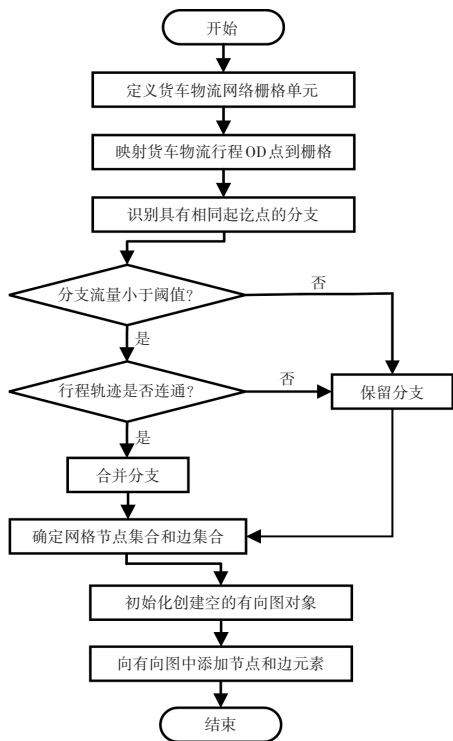


图3 货车物流网络构建流程

2.2.3 网络基序识别

常见的基序识别算法有基于双向扩展搜索策略的Mfinder算法、基于枚举的FANMOD算法等。本文选用Dotmotif算法^[6]，其采用精确子图同构匹配策略，在辨识复杂性较高和变异性较大的网络基序方面具有优势，然而，Dotmotif算法计算效率受限。本文构建的货车物流网络规模较小，适合选用Dotmotif算法。为了识别货车物流网络中的基本结构，提高基序与物流模式特征之间关系的可解释性，降低分类分析难度，只对由二节点和三节点组成的简单基序进行识别，具体基序如图4所示。由于构建的物流网络为有向网络，图4所示基序中不同节点的边之间存在不连通、单向连通和双向连通3种形式，共有15种。

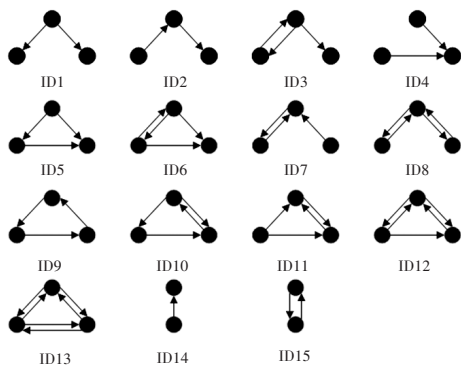


图4 二节点和三节点基序结构汇总

2.2.4 显著性检验

当基序在物流网络中出现的次数远比在随机网络中显著地多时，称该基序在物流网络中的出现是显著的，基序的出现是否显著可以通过统计指标 p 值进行判断。 p 值指原假设为基序出现不显著的前提下，观测到随机网络中该基序出现相较于物流网络中更频繁的概率。如果 p 值小于显著性水平（通常取0.05），则拒绝原假设，可以认为该基序的出现是显著的。 p 值计算公式见式(8)~式(9)：

$$p = \frac{1 + \sum_{i=1}^n C(n)}{1 + n} \quad (8)$$

$$C(n) = \begin{cases} 1, & N_{\text{rand}} \geq N_{\text{real}} \\ 0, & N_{\text{rand}} < N_{\text{real}} \end{cases} \quad (9)$$

式中： n 为随机网络个数， N_{rand} 为基序在某个随机网络出现的次数， N_{real} 为该基序在物流网络中出现的次数。

3 案例验证

3.1 物流模式分类

3.1.1 主成分分析结果

在特征参数降维之前，为评估所构建的多维度特征参数是否适合进行PCA，首先进行KMO（Kaiser-Meyer-Olkin）检验和巴特利特球度检验。计算得KMO值为0.7(>0.5)，巴特利特球度检验显著性水平为0.000(<0.05)，变量间相关性较高，适合进行PCA。其次，使用Scikit-Learn中的PCA方法，通过协方差矩阵对回转半径、日均出行地点、活动熵、日均行驶里程和日均行驶时间5个特征变量进行主成分的提取，并基于特征值计算各主成分方差贡献率及累计贡献率^[17]，结果如表4所示。本文选择特征值大于1的成分为主成分，可以看出，前3个主成分特征值大于1，累计方差贡献率为85.26%，说明第1、第2和第3主成分对量化货车轨迹特征影响显著。因此，本文提取前3个主成分作为聚类算法的输入。

表4 主成分方差贡献率

成分	特征值	方差贡献率/%	累计贡献率/%
第1主成分	4.188	42.33	42.33
第2主成分	2.628	26.56	68.89
第3主成分	1.620	16.37	85.26
第4主成分	0.845	8.54	93.80
第5主成分	0.613	6.20	100.00

3.1.2 轨迹特征聚类结果

现有研究中,货车物流模式通常被划分为2~4类^[2,9]。因此,本文将聚类中心数目 K 分别设置为2、3、4进行K-means聚类,然后再使用轮廓系数法对聚类结果进行分析,聚类结果随着 K 值变化情况如表5所示。可以看出,当 $K=2$ 时,聚类结果分布较为均匀, $K=3$ 时, SC 值最大,表明聚类效果最好,当 $K=4$ 时,4类聚类分布不均匀,聚类效果有所下降。

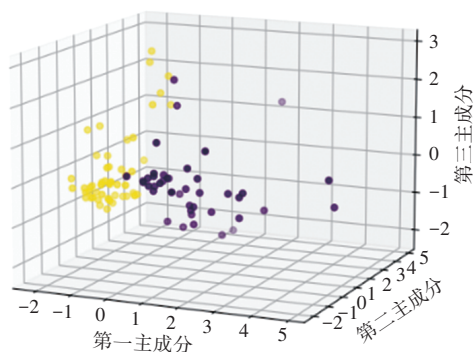
表5 聚类结果汇总

类别	$K=2$	$K=3$	$K=4$
第1聚类	51	49	42
第2聚类	42	33	23
第3聚类		11	19
第4聚类			9
轮廓系数 SC	0.34	0.39	0.37

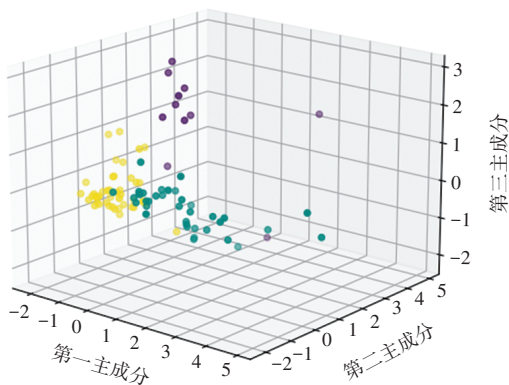
同时采用可视化的方法对上述3类聚类结果进行对比分析,如图5a~图5c所示。可以看出当聚类中心数目为3时,各聚类轮廓清晰、交叠面积较小且样本分布更均衡。综合 SC 轮廓系数和聚类结果可视化图像,选取聚类中心数目 $K=3$ 。

3.1.3 物流模式分类结果

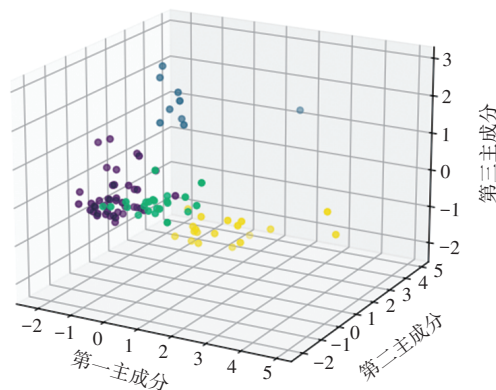
提取出每个类的内部成员原始字段,分别计算每个类的特征参数,结果如表6列所示。



(a) 聚类中心 $K=2$



(b) 聚类中心 $K=3$



(c) 聚类中心 $K=4$

图5 不同聚类中心数目的聚类可视化结果

表6 各个聚类的特征参数汇总

	第1聚类	第2聚类	第3聚类
回转半径/(°)	1.24	0.41	0.23
日均出行地点	4.39	3.57	8.82
活动熵	0.76	0.51	1.14
日均行驶时间/s	25 685.04	14 268.86	20 690.94
日均行驶距离/km	248.83	71.19	100.59
车速平均值/ $\text{km}\cdot\text{h}^{-1}$	67.26	54.80	40.31
发动机平均转速/ $\text{r}\cdot\text{min}^{-1}$	1 892.45	1 705.60	1 461.55
车速标准差/ $\text{km}\cdot\text{h}^{-1}$	29.67	28.59	23.67
加速度标准差/ $\text{m}\cdot\text{s}^{-2}$	0.15	0.18	0.17
发动机转速标准差/ $\text{r}\cdot\text{min}^{-1}$	509.26	517.65	466.93
高速时间比	0.70	0.54	0.34
急加速时间比	0.49	0.35	0.17
聚类数目	11.00	49.00	33.00

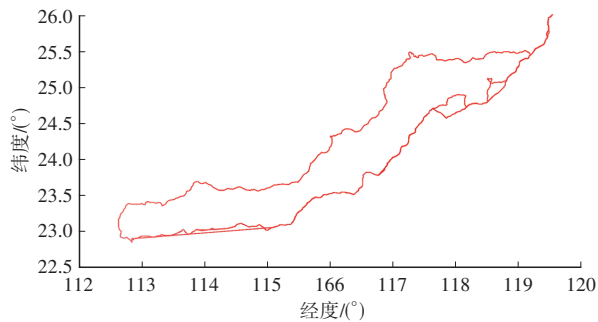
由表6数据可知,3个聚类在轨迹特征上有着显著区别,根据3类货车的轨迹特征将其划分为如下3种物流模式:

(1)第1聚类为长途往返型物流货车。此类货车回转半径、日均行驶时间和日均行驶距离都显著高于其他两类,较高的回转半径表明此类货车活动范围广,较高的日均行驶时间和距离表明此类货车偏向于长途运输。此外,此类货车长时间处于高速行驶状态,与长途运输货车国道和高速公路等主要交通干线的驾驶场景相符合。在物流任务方面,此类货车日均出行地点和活动熵不高,表明其无需频繁装卸货物,主要执行定点送货任务,只在固定的几个货物装卸点中往返。对该聚类中随机抽取的一辆货车进行轨迹数据可视化,如图6a所示,观察到该货车在广州与福州之间进行往返运输,符合长途往返型物流模式定义。

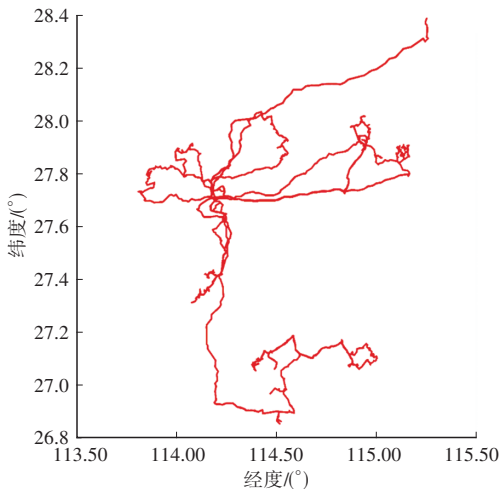
(2)第2聚类为短途定点型物流货车。此类货车的回转半径和日均行驶距离远小于第1类,表明其整体

活动范围较小,属于短途型货车。在物流任务方面,此类货车日均出行地点、活动熵均较低,表明无需频繁装卸货物,主要执行定点送货任务,只在固定的几个货物装卸点中往返。此外,此类货车的平均速度处于中等水平,说明其拥有较为顺畅的行驶条件和较少的停车需求,与定点型物流货车工作场景相符合。对该聚类中随机抽取的一辆货车进行轨迹数据可视化,如图6b所示,观察到该货车在韶关市附近的3个定点间进行往返运输,符合短途定点型物流模式定义。

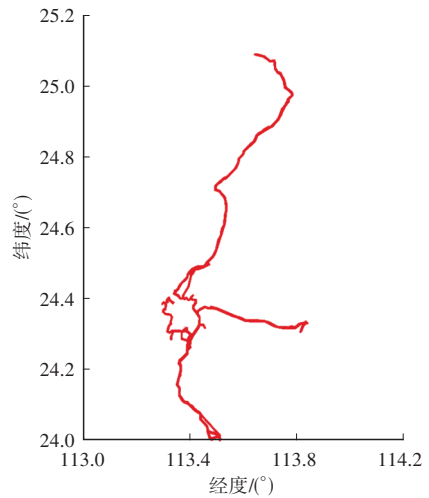
(3)第3聚类为短途多点型物流货车。此类货车的回转半径最小,表明该类货车的运输范围集中在较小的区域,属于短途型货车。但是此类货车日均行驶时间、日均行驶距离、日均出行地点和活动熵的值高于同为短途型的第2聚类。该结果表明其虽然活动范围小,但物流任务的多样性、离散度高,需要执行多点送货任务,在小范围内多个装卸点频繁往返。此外,此类货车平均速度最小,说明此类货车需要频繁停车,整体行驶道路环境为低速环境,反映了短途多点型物流模式特点。对该聚类中随机抽取的一辆货车进行轨迹数据可视化,如图6c所示,观察到该货车在宜春市范围内的网状和环状运输路径,符合短途多点型物流模式定义。



(a)长途往返型货车轨迹



(b)短途多点型货车轨迹



(c)短途定点型货车轨迹

图6 不同物流模式货车的物流轨迹可视化

3.2 物流模式验证

3.2.1 物流网络建立结果

按照前文所述物流网络构建流程,得到了可以有效表示OD数据的有向物流网络,以货车A为例,有向物流网络可视化结果如图7所示。图中圆圈节点代表主要停留点,节点间由有向箭头连接,代表运输方向。

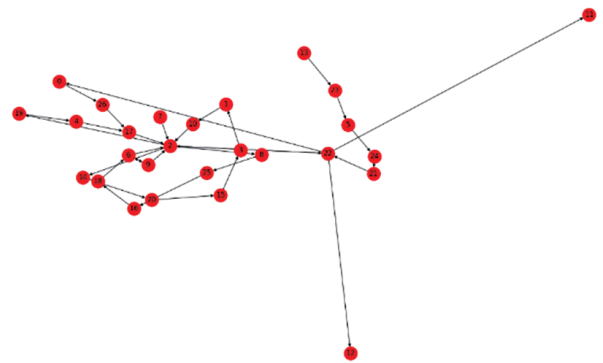


图7 货车A的有向物流网络可视化

3.2.2 基序显著性检验结果

根据基序识别方法,本文对货车OD数据构建的物流网络及生成的随机网络进行基序识别,计算每辆货车的15个不同基序的 p 值,并计算每个聚类内部货车的15个不同基序的 p 值平均值,结果如表7所示。从图表中可以看出,只有个别基序的出现具有显著性,且不同聚类中具有显著性的基序不相同。设定显著性条件为 $p \leq 0.05$,识别每辆货车物流网络中显著性基序,3个聚类的统计结果分别如图7a~图7c所示。

本文使用热图来展示92辆货车各自物流网络中的15个不同基序的显著性。图7横纵轴交叉处颜色表示对应货车物流网络中对应基序的 p 值;星号标记则表示对应货车物流网络中对应基序的 $p < 0.05$,

拒绝原假设,即能认为对应货车物流网络中该基序的出现是显著的。可以观察到,图7a所示的聚类1中基序ID2的 p 值普遍较小,在更多的货车物流网络中具有统计显著性。图7b所示聚类2中的基序ID3、ID7、ID8、ID15以及图7c所示聚类3中的基序ID2、ID3、ID7、ID8、ID9、ID15与之类似。

表7 基序识别显著性检验

motif ID	p 值		
	聚类1	聚类2	聚类3
1	0.46	0.52	0.59
2	0.15	0.55	0.27
3	0.38	0.17	0.13
4	0.59	0.48	0.69
5	0.60	0.43	0.48
6	0.51	0.72	0.54
7	0.29	0.15	0.25
8	0.43	0.09	0.11
9	0.86	0.73	0.07
10	0.35	0.50	0.43
11	0.48	0.68	0.51
12	0.57	0.34	0.36
13	0.78	0.46	0.52
14	0.61	0.43	0.50
15	0.27	0.10	0.21

3.2.3 物流模式验证分析

Gan等^[14]对中国西南片区货车轨迹进行分析,得到3种货车出行链,本文结合货车出行轨迹观测结果将3种出行链分解,得到3种典型的空拓结构。

(1)两点型:货车在装卸点间进行往返运输,例如基序ID15。ID15包含一对双向连接的节点,文献[16]称这类结构为“循环链”,特点是具有往返的路径和非重叠的起始点和终止点。

(2)走廊型:货车从始发点经停中间点,最后到达终止点,例如基序ID2。ID2由单向的“链式”结构组成,起始点和终止点非重叠。

(3)环型:货车沿封闭的环状路线进行货运,例如基序ID9。ID9被称为“单循环”,这类结构同样由单向的“链式”结构组成,但其从起始点最后仍然回到起始位置。

对于划分为长途往返型的聚类1,基序ID2的出现较显著,即走廊型的空拓结构在长途往返型货车的物流网络中更常见。因为长途往返型货车物流始发站与终点站之间距离较远,不会频繁地在两个站

点之间往返,而是会沿着预定路线进行一次完整的往返运输。

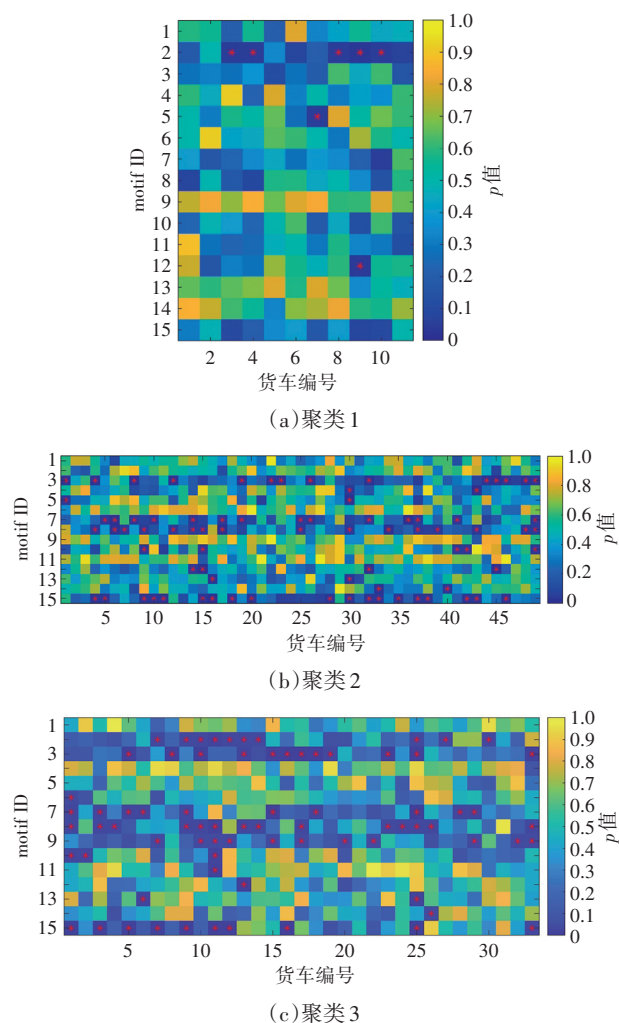


图7 聚类显著性基序识别结果可视化

对于划分为短途定点型的聚类2,基序ID3、ID7、ID8、ID15的出现较显著,此类基序结构中均包含以基序ID15为代表的两点型空拓结构,但同样包含两点型结构的基序ID6、ID10~ID13并不显著,说明短途定点型货车的物流网络中封闭的环形结构不是常见特征。短途定点型货车物流网络通常是以中心点和几个主要的配送点为基础,货车从这些中心点出发将货物运输至配送点,然后返回中心点重新装载。这种物流模式侧重于对固定配送点需求的响应,并且通常涉及大批量的货物装载,因此,在多个配送节点之间构建连续的环形运输路线在这种模式下并非必要,也不符合其运营效率和物流需求。

对于划分为短途多点型的聚类3,基序ID2、ID3、ID7、ID8、ID9、ID15的出现较显著,显著性基序的种类较多,两点型、走廊型、环形的空拓结构在短途多点型货车的物流网络中均常出现。因为短途多点型货

车的物流任务多样性程度高,配送点多和时效性要求高的特点造成配送路线复杂,两点型往返运输路线、走廊型连续运输路线以及环型运输路线普遍存在。

4 结束语

本文由货车GPS数据构建了回转半径、活动熵、日均出行地点、日均行驶距离和日均行驶时间5个特征参数,较全面地量化表征货运活动在时空维度上的特征,进而通过K-means聚类将货车物流模式划分为长途往返型、短途定点型、短途多点型3类。

以货车OD数据构建有向物流网络,从而引入图论中的网络基序识别技术,挖掘了不同物流模式下显著性基序的分布特征。通过从货车出行链提取出两点型、走廊型和环型3种典型拓扑结构,并将其与网络基序相联系,很好地解释了不同物流模式货车物流网络中显著性基序差异。这一方法不仅验证了通过聚类分析得出的物流模式分类结果的准确性,而且为理解物流网络的典型特征和结构提供了新的视角。

根据货车驾驶员物流模式识别结果,网络货运平台能够为不同货运经营者分配其更偏好的货运任务,例如,对于物流模式为长途往返型的货运经营者,平台可优先为其匹配跨区域、点对点的货运任务,对于偏好短途的货运经营者,平台可以优先匹配更多邻近城市或者固定区域内的货运任务。进一步地,通过探究各类型货车的主要物流模式,平台能够针对货运任务的需求更精确地调度不同类型货车,优化运力资源配置,提高运输效能。

参 考 文 献

- [1] ZHANG M, DU X. An Empirical Study of The Road Freight Prices in China [C]// Proceedings of the Asia-Pacific Conference on Intelligent Medical 2018 & International Conference on Transportation and Traffic Engineering 2018, 2018: 155-160.
- [2] 何昱欣, 甘蜜, 钱秋君, 等. 基于出行链视角的城际货车司机出行偏好挖掘方法研究[J]. 工程管理科技前沿, 2024, 43(5): 25-32.
- [3] 覃文文, 鄢祺阳, 谷金晶, 等. 重载货车驾驶人驾驶风格识别与量化研究[J]. 交通运输系统工程与信息, 2022, 22(4): 137-148.
- [4] 关庆锋, 任书良, 姚尧, 等. 耦合手机信令数据和房价数据的城市不同经济水平人群行为活动模式研究[J]. 地球信息科学学报, 2020, 22(1): 100-112.
- [5] MA X, MCCORMACK D E, WANG Y. Processing Commercial Global Positioning System Data to Develop a Web-Based Truck Performance Measures Program[J]. Transportation Research Record, 2011, 2246(1): 92-100.
- [6] DUAN M, QI G, GUAN W, et al. Comprehending and Analyzing Multiday Trip-Chaining Patterns of Freight Vehicles Using a Multiscale Method with Prolonged Trajectory Data [J]. Journal of Transportation Engineering, Part A: Systems, 2020, 146(8): 04020070.
- [7] AKTER T. Truck Activity Pattern Classification Using Anonymous Mobile Sensor Data [D]. Fayetteville: University of Arkansas, 2019.
- [8] 李江涛. 基于货运卡车轨迹大数据的货运OD信息提取方法研究[D]. 北京: 北京交通大学, 2019.
- [9] SIRIPIROTE T, SUMALEE A, HO H. Statistical Estimation of Freight Activity Analytics from Global Positioning System Data of Trucks [J]. Transportation Research Part E, 2020, 140(8): 101986.
- [10] MA X, WANG Y, MCCORMACK E, et al. Understanding Freight Trip-Chaining Behavior Using a Spatial Data-Mining Approach with GPS Data [J]. Transportation Research Record, 2016, 2596(1): 44-54.
- [11] YOU S I, RITCHIE S G. A GPS Data Processing Framework for Analysis of Drayage Truck Tours [J]. KSCE Journal of Civil Engineering, 2018, 22(4): 1454-1465.
- [12] GAN M, NIE Y, LIU X, et al. Whereabouts of Truckers: An Empirical Study of Predictability [J]. Transportation Research Part C, 2019, 104: 184-195.
- [13] 曹劲舟, 涂伟, 李清泉, 等. 基于大规模手机定位数据的群体活动时空特征分析[J]. 地球信息科学学报, 2017, 19(04): 467-474.
- [14] WU L, YANG L, HUANG Z, et al. Inferring Demographics from Human Trajectories and Geographical Context [J]. Computers Environment and Urban Systems, 2019, 77(9): 101368.
- [15] SCHNEIDER C M, BELIK V, COURONNÉ T, et al. Unravelling Daily Human Mobility Motifs [J]. Journal of The Royal Society Interface, 2013, 10(84): 20130246.
- [16] MATELSKY J K, REILLY E P, JOHNSON E C, et al. DotMotif: An Open-Source Tool for Connectome Subgraph Isomorphism Search and Graph Queries [J]. Scientific Reports, 2021, 11(1): 13045.
- [17] 杜宇程, 李会民, 宋尚斌, 等. 基于主成分分析和聚类的营运车辆道路行驶工况构建研究[J]. 汽车科技, 2024(4): 47-54.

(责任编辑 明慧)