文章编号：1003-1421(2026)04-0126-11
引用格式：熊立龙，宋丽英，吴丛铭，等. 基于聚类分析的平陆运河开通对广西铁路货运影响研究[J]. 铁道运输与经济，2026，48(4)：126-136，175.
XIONG Lilong，SONG Liying，WU Congming，et al. Research on Impact of Opening of Pinglu Canal on Railway Freight Transport in Guangxi Based on Cluster Analysis[J]. Railway Transport and Economy，2026，48(4)：126-136，175.

基于聚类分析的平陆运河开通对广西铁路货运影响研究

熊立龙[1,2]，宋丽英[1,2]，吴丛铭[3]，王 斌[4]

(1.北京交通大学 交通运输学院，北京 100044；2.北京交通大学 中国综合交通研究中心，北京 100044；3.广西钦州保税港区宏港码头有限公司 技术部，广西 钦州 535000；4.中国铁道科学研究院集团有限公司 电子计算技术研究所，北京 100081)

摘 要：为探究平陆运河开通对广西铁路货运的影响机制，构建科学的城市受影响模式分类体系，本研究基于改进的Influ-Logit分担率模型，测算广西壮族自治区10个重点城市3类货种在4种运输路径下的分担率，据此构建多维度指标体系；基于灰色关联度分析和随机森林特征重要性评估筛选特征，并结合主成分分析降维和聚类分析对城市进行分类；设计敏感性实验，验证了聚类结果的鲁棒性。研究发现平陆运河开通致使铁路运输分担率呈下降趋势，但不同城市受影响程度存在空间异质性，产业结构是决定铁路运输韧性的核心因素，而大宗散货平陆运河分担率变化量特征重要性最高。聚类结果显示广西城市可划分为2大类，百色市为第1类，具有铁路分担率高、平陆运河替代效应小的特点；其他9个城市为第2类，可细分为4个亚类，其中南宁、贵港、崇左等中部城市因邻近北部湾港口，铁路运输面临最大竞争压力。本研究揭示了产业结构、地理区位与既有运输依赖度是影响铁路货运韧性的关键因素，为铁路部门制定差异化发展策略提供了科学依据。

关键词：平陆运河；铁路货运；运输分担率；随机森林；聚类分析
中图分类号：U2-9 **文献标识码**：A **DOI**：10.16668/j.cnki.issn.1003-1421.20251005002

Research on Impact of Opening of Pinglu Canal on Railway Freight Transport in Guangxi Based on Cluster Analysis

XIONG Lilong[1,2]，SONG Liying[1,2]，WU Congming[3]，WANG Bin[4]

(1. School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; 2. China Comprehensive Transportation Research Center, Beijing Jiaotong University, Beijing 100044, China; 3. Technology Department, Honggang Port Co., Ltd., Guangxi Qinzhou Free Trade Port Area, Qinzhou 535000, Guangxi, China; 4. Institute of Electronic Computing Technology, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China)

Abstract: To explore the impact mechanism of the Pinglu Canal opening on railway freight transport in Guangxi and construct a scientific classification system for urban impact patterns, this paper

收稿日期：2025-10-05 **修订日期**：2026-03-10
通信作者：宋丽英(1978—)，女，北京人，北京交通大学交通运输学院教授。
基金项目：广西科技重大专项(桂科AA23062021-2)

calculated the modal split rates of three cargo types across four transport routes for 10 key cities in Guangxi Zhuang Autonomous Region based on an improved Influ-Logit modal split model and thereby constructed a multi-dimensional indicator system; features were screened based on grey relational analysis and random forest feature importance assessment, and the cities were classified by combining principal component analysis for dimensionality reduction and cluster analysis; sensitivity experiments were designed to validate the robustness of the clustering results. The research results show that the opening of the Pinglu Canal leads to a declining trend in railway transport modal split rates, but spatial heterogeneity exists in the degree of impact across different cities. The industrial structure is the core factor determining railway transport resilience, and the change in the Pinglu Canal modal split rate for bulk cargo has the highest feature importance. The clustering results indicate that cities in Guangxi can be divided into two major categories. Baise City belongs to Category 1, which is characterized by a high railway transport modal split rate and a small substitution effect from the Pinglu Canal; the other nine cities belong to Category 2, which can be subdivided into four subcategories, among which central cities such as Nanning, Guigang, and Chongzuo face the greatest competitive pressure in railway transport due to their proximity to Beibu Gulf ports. This paper reveals that industrial structure, geographical location, and existing transport dependency are the key factors affecting railway freight transport resilience, which provides a scientific basis for the railway departments to formulate differentiated development strategies.

Keywords: Pinglu Canal; Railway Freight Transport; Modal Split Rate; Random Forest; Cluster Analysis

0　引言

平陆运河作为西部陆海新通道建设的重要战略工程，直接连通西江航运干线与北部湾沿海港口，显著缩短了西南地区货物出海距离。运河通航后，将改变以往以西江干线经珠三角港口为核心的江海联运模式，形成经平陆运河由北部湾出海的水路、经西江干线由珠三角出海的水路、由北部湾出海的铁路和公路等多种运输路径并存的新格局，对多式联运结构特别是铁路货运分担率带来深刻影响。这种影响呈现两方面特征，一方面，运河开通完善了北部湾港口集疏运体系，为铁路运输结构优化创造条件；另一方面，水运成本优势可能对铁路运输产生分流效应，特别是对邻近运河、以大宗散货运输为主的城市影响更为明显。由于各城市在地理区位、产业基础等方面存在差异，其铁路货运受影响的程度呈现明显的空间异质性，统一的发展战略难以适用。因此，有必要基于科学的分类方法识别不同城市的受影响模式，为铁路部门制定差异化策略提供依据。

在分担率测算方面，李一萌等[1]构建了公铁竞争市场下的累积前景理论-后悔理论(CPT-RT)旅客出行选择模型，孟祥涛等[2]基于广义费用函数构建了双参数估计的公铁分担率模型，王雪鑫[3]构建了巢式Logit模型以研究城市轨道交通新线开通对交通方式分担的演化影响。而在运河开通分担率测算领域，Li等[4]提出了基于广义运输成本的Attraction-Logit模型预测平陆运河腹地集装箱转移量，Van等[5]分析了巴拿马运河扩建对跨大西洋集装箱运输路径选择的影响，杨晓松等[6]结合生长曲线模型与改进的Influ-Logit模型预测京杭运河复航后的货运分担率变化，胡裕等[7]则采用定性分析与定量预测相结合的方法，预测了长江航运向新建的湘桂运河通道的分流量。特征重要性分析方面，机器学习方法通过数据驱动评估变量对模型预测的贡献，而非机器学习方法主要依靠统计和专家评估，例如层次分析法[8]，通过专家问卷调查和AHP模型计算识别出关键因素。在基于机器学习的特征重要性评估方面，Loyal等[9]提出了结合随机森林核与降维技术的非参数估计器，能够为每个预测点估计出具有明确方向的局部变量重要性度量，向晓蒙等[10]利用随机森林筛选出牛肉产地判别的关键元素，董良骥等[11]采用随机森林模拟水稻生长变量并分析特征重要性，王禹哲[12]利用随机森林排序事故特征重要性以构建预测模型。张鹏羽等[13]采用XGBoost-SHAP模型，对城轨交通客流量影响因素特征重要性进行评估和分析。聚类分析已广泛应用于交通运输领域，杨芳等[14]利用灰色关联和系统聚类分析甘肃省农产品物流运输影响因素，罗莹等[15]基于系统聚类方法分析了公路货运价格与运输距离的关系，熊纯等[16]

通过系统聚类和K-means聚类优化了高铁快运网络节点的划分。针对高维小样本数据，沈运帷等[17]提出了基于动态时间规整和深度模糊聚类的评估方法，Chen等[18]提出了基于插值的改进K-means算法解决稀疏数据聚类问题，这些方法可有效解决小样本聚类问题。

平陆运河建成通航后，广西壮族自治区重点城市经由北部湾港口的进出口货运活动将受到影响，这种影响将直接作用于相关城市的铁路运输格局，本研究旨在探究上述影响的具体表现及其内在作用机制。采用分担率特征来刻画平陆运河开通对不同城市货运格局的影响趋势，并基于城市货物基准分担率特征(即平陆运河开通前的分担率)、分担率变化特征、距离特征和产业结构特征，利用聚类分析方法对广西重点城市进行科学分类，揭示运河开通对不同类型城市铁路运输的差异化影响机制，为铁路部门制定精准应对策略提供理论支撑和实证依据。

1　数据描述与研究方法

1.1　数据描述

本研究聚焦平陆运河开通对广西铁路货运格局的影响，选取广西10个主要城市作为研究对象，包括南宁市、贵港市、梧州市、柳州市、来宾市、百色市、崇左市、桂林市、贺州市与河池市。选取的10个城市均位于平陆运河的辐射范围内，且依赖内河航运实现江海联运，从而确保样本的代表性和研究的针对性。未选取的钦州市、防城港市、北海市等沿海港口城市能够直接通过港口进行进出口贸易，并不直接受平陆运河影响；玉林市虽临近平陆运河，但其通常需要公路短驳至平陆运河经济带，不在运河直接影响范围内。由此，本研究选择10个典型城市以研究广西内陆城市铁路运输格局受平陆运河的影响。通过百度地图平台获取相关数据，各城市不同运输路径的运输距离如表1所示。

对于运输成本，集装箱货物水路、铁路和公路运输成本为0.041元/(t·km)、0.084元/(t·km)和0.35元/(t·km)；大宗散货水路、铁路和公路运输成本为0.041元/(t·km)、0.067元/(t·km)和0.35元/(t·km)。运输时间是指整个运输环节的整体时间，包括装卸、接驳和在途时间。水路、铁路和公路运输的旅行速度取20 km/h，90 km/h和43 km/h，用于计算在途时间。经内河航道由沿海港口出海的水路运输环节中，货物集结与内河码头装船时间取10 h，到达沿海港口的卸船时间取10 h；货物在铁路货场需进行装卸、编组等作业，到达北部湾港口后还需进行卸车等作业，这些作业环节耗时取12 h；公路运输装卸货物时间定为0.5 h，由此计算出运输时间。其中，对进出口贸易货物的运输时间不作细化区分，以简化模型计算，并突出不同运输路径间的竞争关系变化的研究目标。首先，进出口货物装卸流程结构基本一致，虽然具体装卸地点或仓储需求存在差异，但整体运输时间受其影响较小；同时，装卸作业时间波动较大，难以精确量化，统一假设有助于稳定模型估计，保证预测结果的稳健性；并且鉴于本研究侧重于区域铁路货运分担率的宏观影响评估，认为此假设合理且实用。考虑运输的环保性，参考其他研究[6]，对水路、铁路和公路运输的环保性分别取值为1，0.5和0.025。本研究使用数据来源于既有文献和资料，不涉及具体企业运营敏感信息。

表1　各城市不同运输路径的运输距离　km

Tab.1　Transportation distance of different transport routes for each city

重点城市	经平陆运河由北部湾港口出海水路	经西江干线由珠三角港口出海水路	由北部湾港出海铁路	由北部湾港出海公路
南宁市	290	850	100	128
贵港市	280	638	170	223
梧州市	552	366	392	426
柳州市	598	744	300	308
来宾市	572	718	274	241
百色市	658	1 218	320	344
崇左市	497	1 057	238	167
桂林市	893	707	482	445
贺州市	693	463	527	531
河池市	752	898	327	352

1.2　运输分担率测算模型

采用分担率特征来刻画平陆运河开通对不同城市进出口货物运输格局的影响趋势，首先对数据进行标准化处理，货主客户通常追求运输过程中最小化的运输时间和运输费用，故需先将这2个负效用指标进行正向化处理，再对3个指标进行无量纲化处理。在进行标准化处理之后，利用Influ-Logit模型[6]对平陆运河开通前后重点城市进出口贸易不同运输路径的分担率进行测算。该模型能够有效避免超过临界值后的预测结果随着效用值的增大而呈指

数级增长的弊端，计算公式如下。

$$P_k(i)=\frac{\exp\left(U_k(i)/\bar{U}_k\right)}{\sum_{i=1}^{n}\exp\left(U_k(i)/\bar{U}_k\right)} \tag{1}$$

式中：$P_k(i)$为货种k第i种运输路径的占比；$U_k(i)$表示货种k第i种运输路径的效用值；$\bar{U}_k$为货种k运输路径效用值的均值。

效用值及其均值计算公式如下。

$$U_k(i)=\sum_{v=1}^{n}\theta_v^k Y_v^k(i) \tag{2}$$

$$\bar{U}_k=\sum_{i=1}^{n}\frac{U_k(i)}{n} \tag{3}$$

式中：θ_v^k为货种k影响因素v的权重；$Y_v^k(i)$为货种k第i种运输路径的影响因素v；n表示运输路径的数量，条。

考虑对不同货种取不同的权重系数，通过对货主进行调研，得到3类货种的权重系数如表2所示，可以看出，货主对集装箱货物运输时效性的要求相对于散货更高。

1.3　特征体系的构建

为了通过聚类对平陆运河开通对广西重点城市进出口货物的铁路运输影响模式进行识别，需要构建数据样本的特征体系，通过数据搜集和计算，得到样本特征集如表3所示，其中G表示高附加值商品集装箱的相关分担率特征，%；D表示大宗商品集装箱的相关分担率特征，%；S表示大宗散货的相关分担率特征，%；Δ表示对应货种的分担率变化量特征；L表示北部湾港口到对应城市的运输距离，km。

表2　3类货种的权重系数

Tab.2　Weight coefficient of three cargo types

权重系数	高附加值商品集装箱(k=1)	大宗商品集装箱(k=2)	大宗散货(k=3)
运输费用(v=1)	0.393	0.512	0.594
运输时间(v=2)	0.494	0.391	0.317
环保性(v=3)	0.113	0.097	0.089

表3　样本特征集

Tab.3　Sample feature set

城市	南宁市	贵港市	梧州市	柳州市	来宾市	百色市	崇左市	桂林市	贺州市	河池市
ΔG平陆运河/%	25.78	26.20	24.35	24.11	23.77	25.14	24.81	21.97	24.09	23.19
G西江干线/%	25.58	29.18	35.90	29.16	28.72	23.10	23.74	31.93	36.76	27.62
ΔG西江干线/%	−6.52	−7.59	−8.73	−7.06	−6.88	−5.73	−5.87	−7.07	−8.84	−6.49
G铁路运输/%	39.08	38.85	38.26	40.25	38.97	45.19	40.20	41.09	39.72	42.30
ΔG铁路运输/%	−10.13	−10.25	−9.30	−9.65	−9.20	−11.46	−9.99	−8.79	−9.54	−9.66
G公路运输/%	35.34	31.98	25.84	30.60	32.31	31.71	36.06	26.98	23.52	30.08
ΔG公路运输/%	−9.13	−8.36	−6.32	−7.40	−7.69	−7.95	−8.95	−6.11	−5.70	−7.03
ΔD平陆运河/%	26.19	26.79	25.62	25.25	24.77	26.50	25.60	23.73	25.72	24.61
D西江干线/%	27.15	30.77	38.05	31.26	30.57	25.91	25.79	34.71	39.66	30.08
ΔD西江干线/%	−7.03	−8.19	−9.79	−7.88	−7.58	−6.72	−6.54	−8.23	−10.28	−7.41
D铁路运输/%	38.87	38.93	38.68	40.37	38.95	45.24	39.86	41.34	40.11	42.46
ΔD铁路运输/%	−10.25	−10.53	−9.96	−10.23	−9.64	−12.23	−10.26	−9.74	−10.40	−10.44
D公路运输/%	33.97	30.29	23.27	28.37	30.47	28.85	34.35	23.95	20.24	27.46
ΔD公路运输/%	−8.90	−8.06	−5.87	−7.14	−7.55	−7.55	−8.80	−5.75	−5.04	−6.76
ΔS平陆运河/%	26.43	27.10	26.26	25.84	25.31	27.15	26.00	24.62	26.48	25.36
S西江干线/%	28.23	31.79	39.07	32.50	31.68	27.68	27.11	36.07	40.86	31.54
ΔS西江干线/%	−7.39	−8.58	−10.35	−8.38	−8.01	−7.36	−6.98	−8.89	−10.97	−7.98
S铁路运输/%	38.85	39.21	39.55	40.87	39.33	45.66	39.96	42.25	41.23	43.00
ΔS铁路运输/%	−10.35	−10.75	−10.49	−10.65	−9.98	−12.73	−10.47	−10.43	−11.08	−10.98
S公路运输/%	32.92	29.00	21.38	26.64	28.99	26.66	32.92	21.68	17.91	25.46
ΔS公路运输/%	−8.69	−7.77	−5.42	−6.81	−7.32	−7.06	−8.55	−5.30	−4.42	−6.40
L平陆运河/km	290	280	552	598	572	658	497	893	693	752
L西江干线/km	850	638	366	744	718	1 218	1 057	707	463	898
L铁路运输/km	100	170	392	300	274	320	238	482	527	327
L公路运输/km	128	223	426	308	241	344	167	445	531	352
第一产业GDP占比/%	11.31	19.15	7.57	10.46	21.82	16.84	17.25	26.15	19.98	20.33
第二产业GDP占比/%	19.27	30.17	43.00	39.95	38.80	44.00	42.70	18.56	35.11	29.26
第三产业GDP占比/%	69.42	50.68	49.43	49.59	39.38	39.15	40.04	55.29	44.91	50.40

2　基于聚类方法的城市受影响模式类别划分

2.1　数据预处理与特征筛选

鉴于样本规模较小而特征维度较高，在进行聚类之前需要进行特征筛选，采用相关性分析和特征重要性分析相结合的方法，将重要性高和与其他特征关联性低的特征筛选出来。对特征进行基于灰色关联度法的相关性分析，首先对特征进行初值化标准化处理，再计算灰色关联度，分辨系数取0.5，特征灰色关联度相关性矩阵热力图如图1所示。

基于随机森林的特征重要性分析步骤如下。先对特征进行标准化处理，以ΔG铁路运输、ΔD铁路运输、ΔS铁路运输变化量绝对值的累加值最小作为随机森林算法的目标函数，使用均方误差作为不纯度指标，铁路运输分担率变化量越小，说明铁路运输系统在应对新的运输路径开通时的韧性越高。多次实验表明，特征重要性排序结果稳定，重要性排名前12特征的重要性得分及其累积重要性如表4所示，所有特征的重要性排序如图2所示。

特征重要性分析如图3所示，重要性排名第12的特征是ΔG平陆运河，前12个特征的累积重要性已经达到了86.4%。并且产业结构对铁路分担率韧

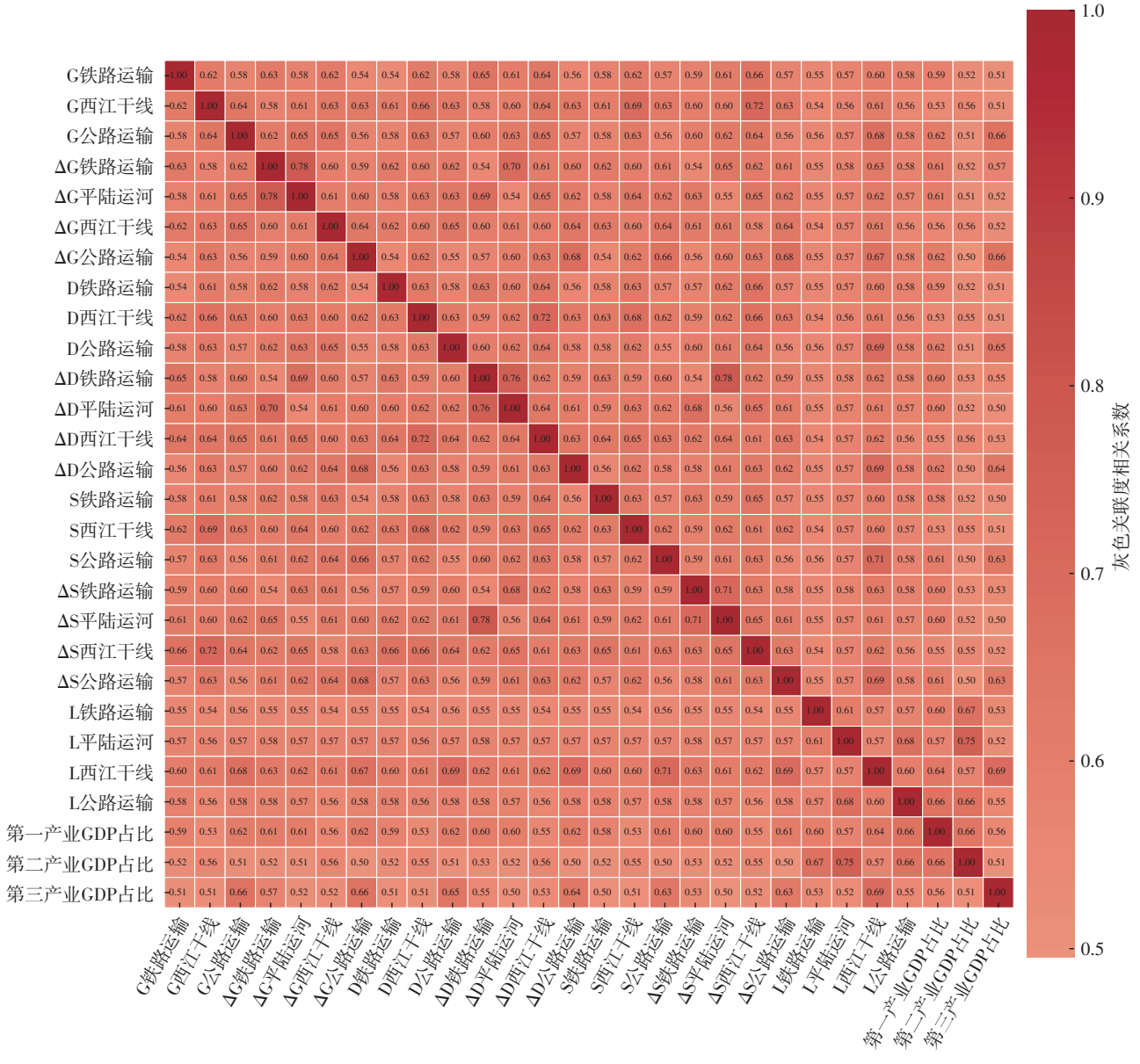


图1　特征灰色关联度相关性矩阵热力图

Fig.1　Heatmap of feature grey relational correlation matrix

表4　重要性排名前12特征的重要性得分及其累积重要性 %

Tab.4　Importance score and its cumulative importance of top 12 features

重要性排名	特征名称	重要性得分	累积重要性
1	ΔS平陆运河	13.61	13.61
2	第三产业GDP占比	13.18	26.78
3	第一产业GDP占比	11.53	38.32
4	G西江干线	10.27	48.59
5	第二产业GDP占比	6.53	55.12
6	G铁路运输	6.27	61.39
7	L西江干线	4.75	66.14
8	S铁路运输	4.64	70.79
9	D铁路运输	4.61	75.39
10	ΔD平陆运河	4.54	79.93
11	ΔG西江干线	3.73	83.66
12	ΔG平陆运河	2.73	86.40

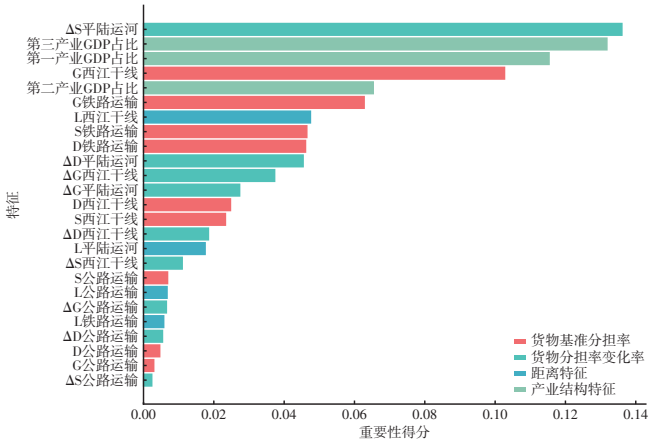


图2　所有特征重要性排序

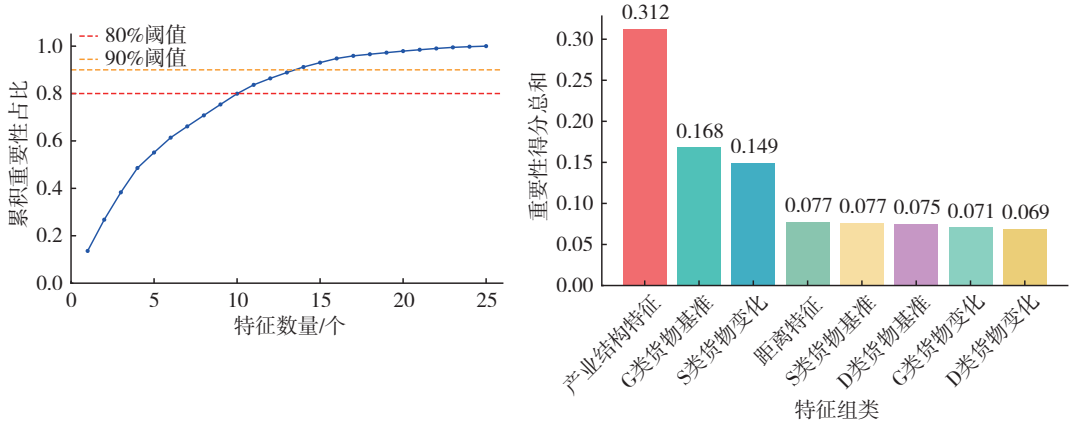
Fig.2　Ranking of all feature importance

性的重要性最高，远高于其他几种特征组。

对特征体系进行筛选，构建出城市受影响聚类特征体系，共有15个特征被考虑作为聚类特征，其累积重要性达到了93.05%。随机森林算法特征重要性计算以ΔG铁路运输、ΔD铁路运输、ΔS铁路运输累加值最小作为目标函数，因此在构建聚类特征体系时，还需要将这3种货类的铁路运输分担率变化量特征考虑在内，筛选得到的最终特征体系及其重要性如表5所示。

2.2　主成分分析

对筛选的特征进行检验，Bartlett检验统计量观测值为768.768，概率P值<0.5，表明数据适合进行因子分析。首先，对筛选得到的特征进行标准化处理，再对特征进行主成分分析(Principal Component Analysis，PCA)，累积方差贡献率阈值设为0.85，得到3个主成分，各主成分方差贡献率从高到低分别为46.62%，21.79%和19.03%，三者的累积方差贡献率达到了87.44%。主成分因子旋转后的因子载荷矩阵热力图如图4所示，根据主成分因子旋转后的因子载荷矩阵，可具体定义主成分因子的类型。主成分PC1在所有的西江干线特征上有较大的载荷，定义为西江干线特征主成分因子，其主要捕获了西江干线运输的地理区位劣势与运输依赖度反向关系。主成分PC2在所有的平陆运河水路运输分担率变化量、部分货种的铁路运输分担率变化量和第一产业GDP占比上有较大的载荷，定义为平陆运河特征主成分因子，其主要捕获了平陆运河开通的替代效应。主成分PC3在所有的铁路运输相关特征和第三产业GDP占比上有较大的载荷，定义为铁路运输特征主成分因子，其主要捕捉了铁路运输市场竞争劣势效应。



a　特征累积重要性曲线

a　Cumulative feature importance curve

b　特征分组重要性分析

b　Feature group importance analysis

图3　特征重要性分析

Fig.3　Feature importance analysis

表5　筛选得到的最终特征体系及其重要性　%

Tab.5　Feature group importance analysis

特征分组	特征名称	重要性得分
S类货物变化	ΔS平陆运河	13.61
产业结构特征	第三产业GDP占比	13.18
产业结构特征	第一产业GDP占比	11.53
G类货物基准	G西江干线	10.27
产业结构特征	第二产业GDP占比	6.53
G类货物基准	G铁路运输	6.27
距离特征	L西江干线	4.75
S类货物基准	S铁路运输	4.64
D类货物基准	D铁路运输	4.61
D类货物变化	ΔD平陆运河	4.54
G类货物变化	ΔG西江干线	3.73
G类货物变化	ΔG平陆运河	2.73
D类货物基准	D西江干线	2.47
S类货物基准	S西江干线	2.33
D类货物变化	ΔD西江干线	1.85
G类货物变化	ΔG铁路运输	—
D类货物变化	ΔD铁路运输	—
S类货物变化	ΔS铁路运输	—

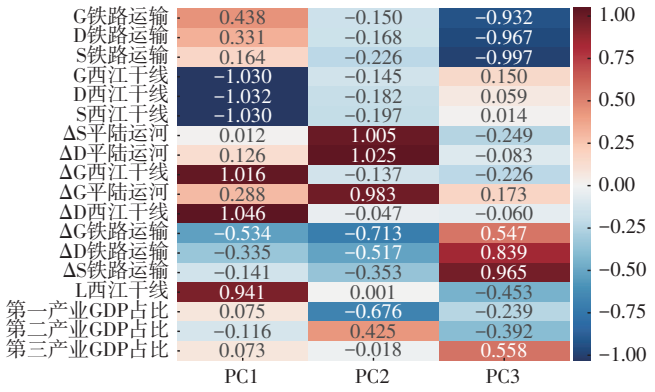


图4　主成分因子旋转后的因子载荷矩阵热力图

Fig.4　Heatmap of factor loading matrix after principal component rotation

2.3　聚类分析与参数优化

采用PCA降维的聚类方法，对城市铁路运输受影响模式类别进行划分。在PCA降维后，首先通过层次聚类法进行预分析，直观展示城市间的层次关系，为后续的K-means方法提供初步聚类数K的参考，PCA降维后城市层次聚类树状图如图5所示，最优簇数为2~4。

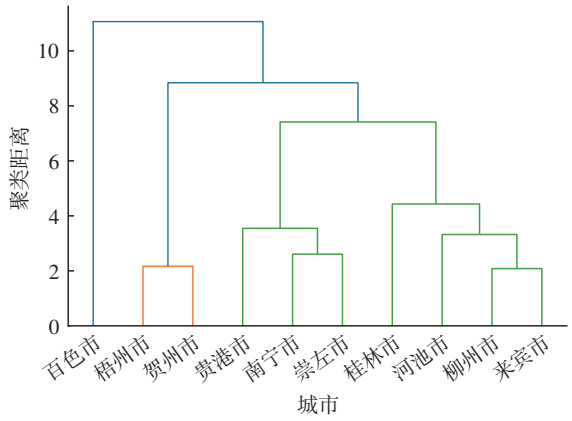
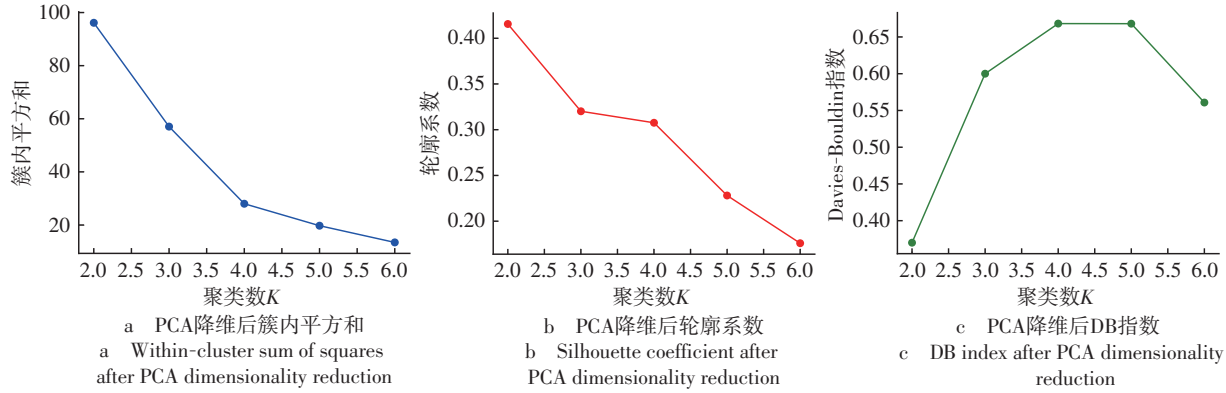


图5　PCA降维后城市层次聚类树状图

Fig.5　Dendrogram of city's hierarchical clustering after PCA dimensionality reduction

基于层次聚类法的簇数量，选取K-means聚类方法K值的大致范围为2~6，采用K-means聚类方法，分别计算不同K值簇内平方和、轮廓系数和Davies-Bouldin指数等指标，PCA降维后聚类指标如图6所示，根据综合结果可以确定最优聚类数为2。经过多次聚类验证，聚类结果对初始中心的选取具有鲁棒性。

对不同K值进行层次聚类，计算与同K值下K-means聚类结果的一致性指标，分别计算调整兰德指数(ARI)、标准化互信息(NMI)，PCA降维后2种聚类结果一致性检验如图7所示。当K=2，3，5，6



a　PCA降维后簇内平方和

a　Within-cluster sum of squares after PCA dimensionality reduction

b　PCA降维后轮廓系数

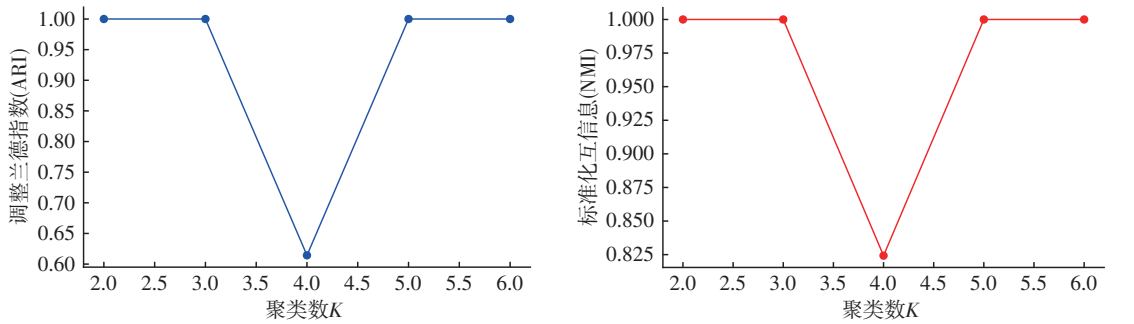
b　Silhouette coefficient after PCA dimensionality reduction

c　PCA降维后DB指数

c　DB index after PCA dimensionality reduction

图6　PCA降维后聚类指标

Fig.6　Clustering metrics after PCA dimensionality reduction



a　PCA降维聚类ARI指标　　b　PCA降维聚类NMI指标

a　ARI metric for PCA dimensionality reduction clustering　　b　NMI metric for PCA dimensionality reduction clustering

图7　PCA降维后2种聚类结果一致性检验

Fig.7　Consistency test of two clustering results after PCA dimensionality reduction

时，2种聚类方式的一致性指标计算结果为1，聚类结果一致，因此最终最优聚类数为2。

基于分析的结果，当K=2时，聚类的质量最优，且一致性水平较高，PCA降维后三维聚类结果可视化如图8所示，样本被划分为了2类，第1类只有百色市，第2类为其他9个城市。

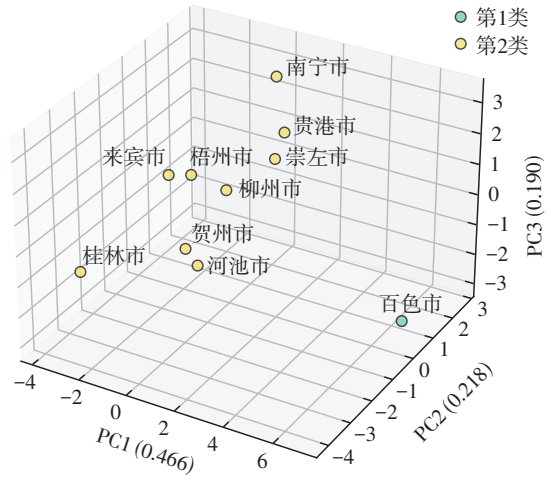


图8　PCA降维后三维聚类结果可视化

Fig.8　Visualization of three-dimensional clustering results after PCA dimensionality reduction

PCA降维变换空间的聚类中心如表6所示，簇0的聚类中心在PC1维度上为正数且较大，在PC2维度上为正数且接近于0，而在PC3维度上为负数且绝对值较大；簇1的聚类中心在PC1和PC2维度上为负数且接近于0，在PC3维度上为正数且接近于0。

2.4　聚类效果的对比实验

围绕PCA降维进行消融对比实验，对标准化处理后的原始数据直接进行聚类分析。基于原始数据的聚类分析实验除了没有进行PCA降维处理之外，

表6　PCA降维变换空间的聚类中心

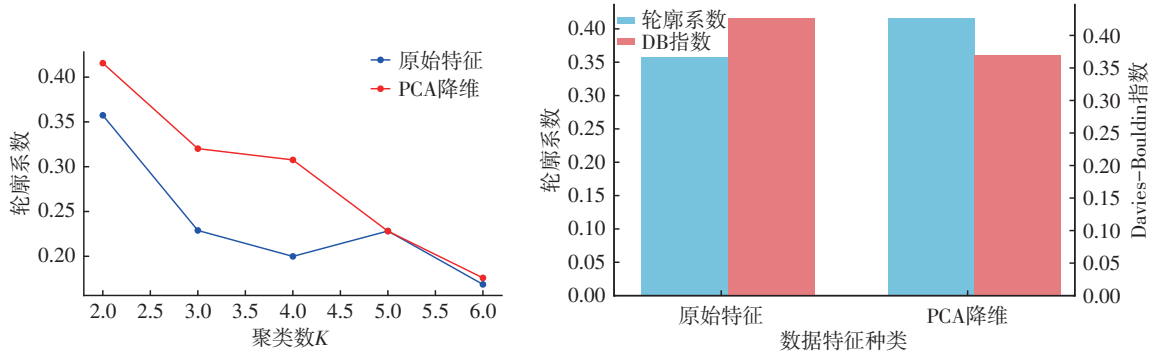
Tab.6　Clustering centers in PCA dimensionality reduction transformation space

簇	类别	PC1	PC2	PC3
簇0聚类中心	第1类聚类中心	6.947	0.341	−2.596
簇1聚类中心	第2类聚类中心	−0.772	−0.038	0.288

其他实验过程与上文的PCA降维聚类方法完全一致，根据各项指标，得到最终最优聚类数为2，并且基于原始数据的分类结果与PCA降维后的分类结果一致。通过实验，原始数据与PCA降维聚类质量对比如图9所示，蓝色表示原始数据聚类质量评估指标，红色表示PCA降维后的聚类质量评估指标，可以得出PCA降维后的聚类结果质量相对较高，且2种方法的分类结果是一致的。因此，以图8所示的聚类结果为最终城市分类结果，即样本被划分为了2类，第1类只有百色市，而第2类为其他9个城市。

2.5　敏感性分析实验

在主成分分析中，累积方差贡献率阈值设为0.85是基于经验的取值，为了验证聚类效果的鲁棒性，补充对该阈值进行敏感性分析实验，累积方差贡献率阈值敏感性分析实验结果如表7所示。当该阈值取值范围为75%~87%时，主成分数保持为3；当阈值取值为90%时则增加至4个主成分，新增的主成分为PC4，累积方差贡献率为95.75%。聚类实验结果显示，阈值在一定范围内调整，最优聚类结果均保持一致，表明该方法对阈值选择具有较强鲁棒性。综上，累积方差贡献率阈值的适量调整不会显著影响聚类分析效果，保证了方法的稳定性和科学性。



a　原始数据与PCA降维聚类质量优化曲线
a　Clustering quality optimization curves for original data and PCA dimensionality reduction

b　原始数据与PCA降维聚类最优K值聚类指标对比
b　Comparison of clustering metrics at optimal K values between original data and PCA dimensionality reduction

图9　原始数据与PCA降维聚类质量对比

Fig.9　Comparison of clustering quality between original data and PCA dimensionality reduction

表7　累积方差贡献率阈值敏感性分析实验结果

Tab.7　Experimental results of sensitivity analysis on cumulative variance contribution rate threshold

阈值/%	备注	主成分数	累积方差贡献率/%	结论
75	降低10%	3	87.44	实验结果完全一致
80	降低5%	3	87.44	实验结果完全一致
83	降低2%	3	87.44	实验结果完全一致
85	基准值	3	87.44	基准情况
87	增加2%	3	87.44	实验结果完全一致
90	增加5%	4	95.75	增加一个主成分，最优聚类结果一致

3　实验结果分析

3.1　特征重要性分析

基于随机森林算法的特征重要性评估结果显示，以各货种铁路运输变化量绝对值累加值最小为目标，特征重要性呈现3个层次。高重要性特征累积重要性达到了48.59%，包括ΔS平陆运河、第三产业GDP占比、第一产业GDP占比和G西江干线；中等重要性特征累积重要性大约为50%~85%，包括第二产业GDP占比、G铁路运输、L西江干线等；低重要性特征主要为公路运输的相关指标。分析揭示3项关键规律：其一，产业结构是核心影响因素，产业GDP占比累积重要性达31.24%，表明产业结构是决定铁路运输韧性的核心要素；其二，大宗货物对低成本水路新通道响应敏感，大宗散货和大宗商品集装箱的平陆运河分担率变化量重要性显著，反映了大宗货种更易受水运成本优势吸引；其三，既有运输结构影响韧性，铁路运输基准分担率的高重要性说明城市原有的铁路依赖度是系统韧性的重要基础。公路运输特征重要性较低，表明在长距离、大宗货物运输中公路并非主要竞争对手。

3.2　基于聚类的城市分类分析

聚类方法将广西10个城市划分为2大类。第1类仅包含百色市，其PCA空间得分为(6.947，0.341，−2.596)，表明百色市距西江干线较远，原有西江分担率较低、铁路分担率较高，公路运输分担率为中等水平，平陆运河开通产生了一定的替代效应，但铁路运输仍具有相对优势，其作为云南、贵州等西南省份经广西出海的重要门户，在支撑川渝黔地区货物向北部湾港口转运中仍承担着重要作用，平陆运河的开通对其运输通道功能提出了更高要求，有利于提升其多式联运和区域联通能力。

第2类包含其他9个城市，聚类中心接近原点，代表性相对有限，内部呈现显著异质性。为进一步揭示城市间的差异化特征，对第2类城市进行进一步细分析，根据K=5的聚类结果，将第2类城市分为4组。第1组为桂林市，得分为(−2.301，−4.049，−0.275)，该城市位于广西东北部，距北部湾相对较远，西江干线运输具备一定的优势，原有公路运输分担率相对较低，平陆运河开通的替代效应相对有限，其主要为川渝黔及湘南地区货物流向北部湾和珠三角的传输节点，具备较为稳定的铁路运输基础。第2组为来宾市、柳州市和河池市，得分为(−0.206，−1.175，0.394)，该组城市位于广西北部，距离北部湾相对较远，原有公路运输分担率为中等水平，平陆运河具有一定的替代效应但是相对较弱，这些城市铁路运输主要连接贵州、四川、重庆

等西南地区，并且其铁路运输市场承受着较小的来自西江干线的竞争压力。第3组为梧州市和贺州市，得分为(−3.344，1.628，−2.121)，该组城市位于广西东部，距珠三角港口较近，原有西江干线运输分担率较高，公路运输分担率较低，平陆运河开通对其产生显著的替代效应，通过分流原本流向珠三角的货物，能够有效缓解珠三角港口的货运压力，提升区域水路网络的联动和运输效率，其原有铁路运输分担率较高，但在平陆运河开通后面临较大的竞争压力，铁路运输韧性相对较弱。第4组为南宁市、贵港市和崇左市，得分为(0.887，1.326，1.977)，该组城市位于广西中部，邻近北部湾港口，西江干线运输区位呈现一定劣势，平陆运河开通对其产生显著的替代效应，其原有公路运输分担率明显高于平均水平，公路与铁路之间已存在一定的竞争关系，平陆运河开通后将进一步压缩铁路货运的市场空间。不过，南宁市作为中国−东盟自由贸易区的重要枢纽，在打造国际物流中心的同时积极推进多式联运和西部陆海新通道班列的发展，为铁路货运市场提供了另一条发展路径，铁路货运若能融入多式联运体系，将竞争关系转化为分工协作，不仅可以稳住货运班列这一优势市场，也有助于降低单一运输方式的替代风险，推动该组城市货运结构向更具韧性的多式联运模式演进。

总体而言，平陆运河的开通促使广西货运市场在运输路径和模式上发生结构性调整，水路运输比例整体提升，对传统铁路运输构成梯度性的替代和竞争压力，不同城市因地理位置和产业特点，其铁路运输的韧性存在一定差异，未来应结合城市定位，强化多式联运枢纽建设，充分发挥水路与铁路的互补优势，推动省际及跨境货运协同发展，进而提升广西区域物流系统的整体效率和竞争力。

3.3　主成分阈值敏感性实验分析

根据敏感性分析实验结果，若将累积方差贡献率阈值设为0.9，主成分数则增加为4，各主成分方差贡献率从高到低分别为46.62%，21.79%，19.03%和8.31%，四者的累积方差贡献率达到了95.75%。并且根据主成分因子旋转后的因子载荷矩阵，前3个主成分的载荷表现并没有明显的变化，三者所表征的特征仍然依次为西江干线特征主成分因子、平陆运河特征主成分因子和铁路运输特征主成分因子，新增的主成分4(PC4)则仅在第二产业GDP占比和第三产业GDP占比有较大的载荷，定义为产业结构特征主成分因子。最优聚类结果与基准实验(阈值设为85%)一致，将10个城市划分为2类，第1类仅包含百色市，聚类中心为(6.947，0.341，−2.596，0.146)，第2类为其他9个城市，聚类中心为(−0.772，−0.038，0.288，−0.016)，两者在PC4上的得分均较小。根据聚类结果评估指标，对第2类城市进行进一步细分析，根据K=5的聚类结果，将第2类城市分为4组，第1组为桂林市、河池市，第2组为来宾市、柳州市和崇左市，第3组为梧州市和贺州市，第4组为南宁市、贵港市，这与基准实验的细化分析聚类结果并不完全一致，但是各组类的PC得分与基准实验的相比，趋势变化并不明显。并且，根据85%和90%阈值的聚类分析结果，来宾与柳州、梧州与贺州、南宁与贵港这3组城市始终归属相同的类别，表明它们在内在特征上存在相似性，反映了这些城市在物流系统中的内在联系和相近的功能定位，该稳定的分类为后续针对不同城市群体制定差异化运输发展和政策支持策略提供了坚实的分类基础和理论支撑。

3.4　面向铁路运输的差异化发展策略

对于第1类城市百色市，应强化多式联运枢纽功能，加强与西南地区铁路货运联系，依托铝工业基地发展铝土矿、氧化铝等重载运输产品。对于第2类第1组城市桂林市，应发挥西江干线优势，重点发展电子产品、精密仪器等高附加值商品铁路运输，加强开展与湖南、贵州等地的货运合作。对于第2类第2组城市柳州市、来宾市和河池市，应依托工业基础，构建产业链协同的铁路货运网络，提供定制化运输服务，并给予基础建设政策倾斜。对于第2类第3组城市梧州市和贺州市，应重点发展铁水联运，将竞争关系转为合作关系，通过差异化市场定位形成优势互补。对于第2类第4组城市南宁市、贵港市和崇左市，应加强建设国际铁路物流港，发展西部陆海新通道班列等国际货运业务，并积极推动多式联运“一单制”。总体而言，第1类城市建议加强基础建设和区域货运合作，而第2类城市建议推动多式联运发展。

4　结论

本研究旨在探究平陆运河开通对广西壮族自治区重点城市从北部湾港口进出口货物的铁路运输格局的潜在影响与作用机制，基于广西10个主要城市的样本数据，通过构建Influ-Logit分担率模型与综合聚类分析方法，量化分析了运河开通前后不同货种、不同路径的运输分担率变化，并据此对城市受影响模式进行了科学分类与机理阐释，得到如下结论。

（1）平陆运河的开通将显著影响区域运输格局，其分流效应对铁路货运产生了双重影响，这种影响具有显著的空间异质性。

（2）城市产业结构、既有运输结构依赖度及与运河港口的地理区位是决定其铁路货运受影响程度的关键因素。

（3）通过聚类分析将广西城市划分为2大类别，并对第2类细分4个亚类，刻画了不同城市的受影响的模式特征。

基于上述结论能够为铁路部门应对运河开通制定差异化的发展策略，但是本研究仍存在以下不足。一方面，模型中对运输费用和时间的测算主要用于刻画分担率变化趋势，未来可引入实际运营数据进行校准，以提升预测精度；另一方面，研究主要聚焦于区域内的运输，未来可进一步将研究范围拓展至中国西南地区长距离运输，分析运河在更大尺度上对铁路联运的影响。

参考文献：

[1] 李一萌，单杏花，韩慧婷，等．高速铁路旅游计次票产品定价优化研究[J]．铁道运输与经济，2025，47(6)：137-145，161.
LI Yimeng，SHAN Xinghua，HAN Huiting，et al. Research on Optimization of Product Pricing for High Speed Railway Travel Prepaid Tickets[J]. Railway Transport and Economy，2025，47(6)：137-145，161.

[2] 孟祥涛，张　戎，刘峻麟，等．基于广义费用模型的公铁分担率研究[J]．铁道运输与经济，2025，47(5)：115-122.
MENG Xiangtao，ZHANG Rong，LIU Junlin，et al. Road and Rail Sharing Rates Based on Generalized Cost Model[J]. Railway Transport and Economy，2025，47(5)：115-122.

[3] 王雪鑫．新建城市轨道交通对交通方式分担的演化仿真研究[J]．铁道运输与经济，2021，43(10)：125-132.
WANG Xuexin. Simulation Research on the Evolution of New Urban Rail Transit in Traffic Mode Sharing[J]. Railway Transport and Economy，2021，43(10)：125-132.

[4] LI J S，YANG J Q. Research on Container Demand Analysis and Volume Forecasting in the Hinterlands of Pinglu Canal[M]// Proceedings of the 2024 7th International Symposium on Traffic Transportation and Civil Architecture (ISTTCA 2024). Dordrecht：Atlantis Press International BV，2024：400-408.

[5] VAN HASSEL E，MEERSMAN H，VAN DE VOORDE E，et al. The Impact of the Expanded Panama Canal on Port Range Choice for Cargo Flows from the U. S. to Europe[J]. Maritime Policy & Management，2021，48(5)：610-628.

[6] 杨晓松，范海文，张　楠，等．京杭运河聊城—德州段复航货运量预测研究[J]．山东科学，2022，35(4)：116-125.
YANG Xiaosong，FAN Haiwen，ZHANG Nan，et al. Prediction of Traffic Volume in the Liaocheng-Dezhou Section of the Beijing-Hangzhou Canal after Shipping Resumption[J]. Shandong Science，2022，35(4)：116-125.

[7] 胡　裕，姜丰怡，彭书华，等．湘桂运河建设对长江航运物流格局的影响研究[J]．水运工程，2024(7)：145-149.
HU Yu，JIANG Fengyi，PENG Shuhua，et al. Impact of Xianggui Canal Construction on Logistics Pattern of Yangtze River Shipping[J]. Port & Waterway Engineering，2024(7)：145-149.

[8] 朱建伟．基于层次分析法的建筑塔机事故因素重要性排序分析[J]．建筑机械，2025(8)：135-138.
ZHU Jianwei. Analysis of the Importance Ranking of Factors in Tower Crane Accidents Based on AHP[J]. Construction Machinery，2025(8)：135-138.

[9] LOYAL J D，ZHU R Q，CUI Y F，et al. Dimension Reduction Forests：Local Variable Importance Using Structured Random Forests[J]. Journal of Computational and Graphical Statistics，2022，31(4)：1104-1113.

[10] 向晓蒙，赵超敏，陈念念，等．基于机器学习和元素指纹的牛肉产地溯源[J/OL]．现代食品科技，1-11[2025-10-02]. https：//doi. org/10. 13982/j. mfst. 1673-9078. 2026. 10. 1167.

[11] 董良骥，王雨禾，刘晓俊，等．基于随机森林模型的水稻生长模拟及特征重要性分析[J]．节水灌溉，2025(11)：32-39.
DONG Liangji，WANG Yuhe，LIU Xiaojun，et al. Research on Simulation and Evaluation of Rice Growth Based on Random Forest Model[J]. Water Saving Irrigation，2025(11)：32-39.

[12] 王禹哲．基于改进贝叶斯网络的高速公路交通事故预测方法探究[J]．中国交通信息化，2025(7)：85-87.

（下转第175页）