文章编号：1003-1421(2026)02-0173-08
引用格式：马睿博，汤银英，左鹏辉，等.基于多维特征与聚类分析的铁路物流客户群画像研究[J].铁道运输与经济，2026，48(2)：173-180，204.
MA Ruibo，TANG Yinying，ZUO Penghui，et al. Customer Portrait of Railway Logistics Based on Multi-Dimensional Features and Clustering Analysis[J]. Railway Transport and Economy，2026，48(2)：173-180，204.

基于多维特征与聚类分析的铁路物流客户群画像研究

马睿博，汤银英，左鹏辉，郭赫臣

(西南交通大学 交通运输与物流学院，四川 成都 611756)

摘 要：在“交通强国”和“双碳”战略背景下，针对铁路物流客户画像构建中传统K-means算法存在的特征单一、初始中心敏感等问题，创新性地融合多维特征嵌入与算法优化技术。通过构建动静结合的特征体系，采用肘部法则与轮廓系数优化聚类参数，结合多个评估指标，显著提升了客户分群的准确性，最后应用词云工具等进行客户群特征的可视化呈现。研究成果为铁路物流企业实现精准运力配置和差异化服务提供了有效解决方案，对提升行业数字化运营水平具有重要实践价值。

关键词：铁路物流；客户画像；多维特征；肘部法则；轮廓系数
中图分类号：U294.1+5 **文献标识码**：A **DOI**：10.16668/j.cnki.issn.1003-1421.20250704001

Customer Portrait of Railway Logistics Based on Multi-Dimensional Features and Clustering Analysis

MA Ruibo，TANG Yinying，ZUO Penghui，GUO Hechen

(School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, Sichuan, China)

Abstract: Under the strategic context of building China into a country with strong transportation network and carbon peak and carbon neutrality, in response to the single-feature limitation and initial center sensitivity in traditional K-means algorithm applications for railway logistics customer portrait construction, the multi-dimensional feature embedding and algorithmic optimization techniques were innovatively integrated. By establishing a combined dynamic-static feature system, the clustering parameters were optimized using the elbow method and the silhouette coefficient. Combining multiple evaluation metrics, the accuracy of customer grouping was significantly improved. Additionally, tools such as word clouds were applied to visualize customer cluster characteristics. The results provide an effective solution for railway logistics enterprises to achieve precise capacity allocation and differentiated services, offering important practical value for enhancing the industry's digital operation level.

Keywords: Railway Logistics; Customer Portrait; Multi-Dimensional Features; Elbow Method; Silhouette Coefficient

收稿日期：2025-07-04
通信作者：汤银英(1978—)，女，河南商丘人，西南交通大学交通运输与物流学院副教授。
基金项目：中国铁路北京局集团有限公司科研项目(2025BH01)

在“交通强国”和“双碳”目标推动下，我国铁路物流快速发展，2024年货运量达51.75亿t，同比增长2.8%，但国内研究在铁路物流客户画像上存在特征维度单一、传统K-means算法聚类结果稳定性差、评估体系与业务结合不紧密等问题，难以解决客户需求响应滞后等问题；全球铁路物流市场2024年价值4 168亿美元，预计2025—2034年复合年增长率5.4%，国外研究虽起步早、注重多维度数据融合，却在算法优化上未充分考虑铁路物流业务特殊性，特征嵌入技术与聚类算法结合探索不足，画像体系适配性有提升空间，鉴于此，本研究融合多维特征嵌入技术与优化后的K-means聚类算法，构建更精准的铁路物流客户画像体系，助力我国铁路物流精细化运营和服务升级，提升全球竞争力[1]。

1　铁路物流客户动静态特征分析

铁路物流客户在多个维度存在运输需求的差异性如运输时效要求不同、不同客户采用不同服务模式、不同货物所需的运输要求不同、偏远地区客户等。为更精准地服务铁路物流客户，需要深入了解铁路物流客户静态、动态需求特征[2]。对客户特征进行了多维度的细化分析，重点聚焦于静态特征与动态特征的指标结构，客户特征示意图如图1所示。这一动静态特征分析框架，为后续客户画像模型的构建以及聚类分析奠定了坚实的结构化基础[3-4]。

（1）月均发运频次波动率。是衡量客户月度发运频次稳定性的核心指标，反映客户运输需求的规律性或随机性[5]。

$$\text{月均发运频次波动率} = \frac{\text{月度发运频次的标准差}}{\text{月度发运频次的均值}} \times 100\% \tag{1}$$

（2）运输频次月环比增长率。是指在相邻2个时间周期内，铁路物流客户的运输频次变化的百分比。其计算公式为[6]

$$\text{运输频次月环比增长率} = \frac{\text{本月运输频次} - \text{上月运输频次}}{\text{上月运输频次}} \times 100\% \tag{2}$$

（3）路径成本敏感度。计算货主选择低成本路线与高时效路线的比例变化。其计算公式为

$$\text{低成本线路选择比例} = \frac{\text{低成本路线选择次数}}{\text{总运输次数}} \times 100\% \tag{3}$$

$$\text{高时效路线选择比例} = \frac{\text{高时效路线选择次数}}{\text{总运输次数}} \times 100\% \tag{4}$$

$$\text{比例变化率} = \frac{\text{本期低成本路线选择比例} - \text{上期低成本路线选择比例}}{\text{上期低成本路线选择比例}} \times 100\% \tag{5}$$

（4）服务半径依赖度波动。本研究主要综合衡量客户对某区域的运输依赖程度及其稳定性，结合运输量集中度和波动率，反映客户在区域内的需求集中性及抗风险能力。计算客户在相邻2个时间段内某区域运输量占比的变化率，计算公式为

$$\text{运输量占比} = \frac{\text{区域运输量}}{\text{总运输量}} \times 100\% \tag{6}$$

$$\text{波动率} = \frac{\text{本期区域运输量占比} - \text{上期区域运输量占比}}{\text{上期区域运输量占比}} \times 100\% \tag{7}$$

（5）联运节点扩展速率。是衡量铁路物流客户在多式联运网络中动态调整节点能力的指标，反映客户在一定周期内新增或淘汰节点的净变化速度，体现其网络布局的灵活性。其计算公式为

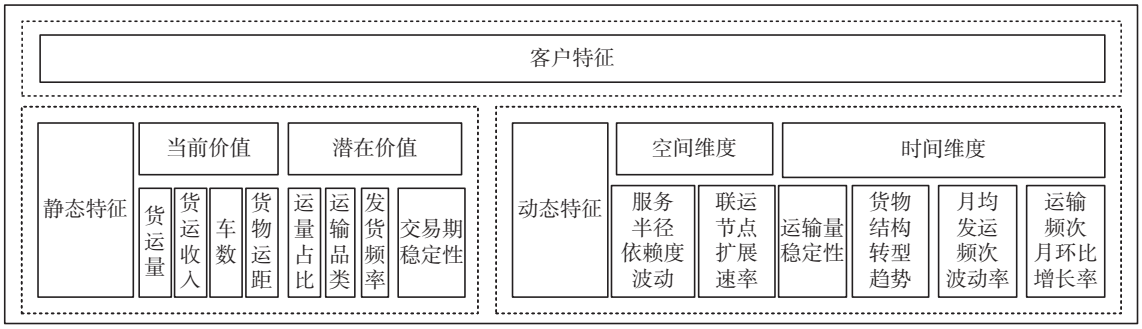

图1　客户特征示意图

Fig.1　Customer characteristics

$$节点扩展速率 = \frac{本期新增节点数 - 本期淘汰节点数}{上期节点总数} \times 100\% \quad (8)$$

2　铁路物流客户画像模型与聚类分析

针对铁路物流客户多维特征所存在的差异性，建立客户画像模型并对具有相似特征的客户群体进行分类。此举突破传统单一维度分析的局限性，将客户信息通过画像模型整合后进行聚类分析，从而将客户划分为不同群体[7]。结合画像与聚类结果，可精准刻画各类客户的需求、行为与偏好，深入挖掘潜在需求，实现市场精准细分，并为不同群体制定差异化服务策略，进而优化资源配置，提升服务质量与客户满意度，从而增强铁路物流企业在市场竞争中的核心优势，实现可持续发展[8]。

2.1　客户画像模型构建

2.1.1　数据获取与预处理

客户画像的构建始于多源数据整合，本研究采集的铁路物流客户属性数据主要源自A铁路局集团公司货运数据库，共包含164 200条发货记录。这些数据涵盖客户静态属性与动态行为特征，为构建多维度客户画像奠定了数据基础。经脱敏处理后，原始数据结构A铁路局集团公司2024年货物运输样表如表1所示[6]。

（1）缺失值处理。针对数据中的缺失值，根据变量的分布特性和重要性采用不同的填充方法。由于数据来源于A铁路局集团公司的数据库，数据可靠且大多为离散型，缺失率也普遍较低，因此本研究采用统计量填充的方法，即对于均匀分布的数据用均值填补缺失，对于倾斜分布的数据采用中位数进行填补。

（2）异常值检测与处理。异常值是数据分布中的离群点，可能由特定业务运营动作或数据本身分布异常导致。本研究采用简单统计分析，结合箱线图和各分位点判断是否存在异常值，并对检测到的异常值进行合理处理，以确保数据的准确性。

在获得的数据表中，利用SPSS26.0对数据进行分析，剔除客户编号、需求单号、发货日期、货运收入为空的数据；剔除重复的、异常的数据。每条交易数据的单号都是唯一的，通过单号信息对交易数据进行统计，找到重复的交易信息。最后对客户数据进行清理统计得出86位客户[9]。

（3）标准化处理。在标准化处理数据时，对月均发运频次波动率和区域依赖波动等进行Z-Score标准化，消除与静态特征的量纲差异。

2.1.2　特征嵌入与标签划分

特征提取是构建客户画像的核心步骤之一。在本研究中，结合客户特征分析及客户实际业务需求，从静态和动态2个维度提取客户特征[10]。

根据上述分析可以构建出基础的客户画像特征体系，客户画像特征模型如图2所示。

（1）静态特征提取。静态特征涵盖客户在较长时间内保持相对稳定的属性，包括交易期稳定性、发货时间间隔以及合作属性。这些特征能够反映客户的身份特征、业务结构和合作基础等核心信息。

（2）动态特征提取。动态特征则强调客户在特定时间窗口内因环境变化、事件触发、交互反馈而产生的时序性、波动性、响应性行为模式，包括时间维度特征和空间维度特征。这些特征能够揭示客户运输需求的规律性和可预测性，以及客户对运输网络和路径的偏好[11]。

采用主成分分析法对客户的各类特征进行分析，以此来验证现有的特征指标是否需要进一步优化。通过Varimax准则对各指标进行分析。

$$V = \sum_{k=1}^{m}\left(\frac{1}{p}\sum_{i=1}^{p}b_{ik}^{4} - \left(\frac{1}{p}\sum_{i=1}^{p}b_{ik}^{2}\right)^{2}\right) \quad (9)$$

式中：b_{ik}为标准化后的因子载荷；p为变量个数，个；m为因子个数，个。

通过特征分析，优化后的客户画像特征模型如

表1　A铁路局集团公司2024年货物运输样表

Tab.1　Sample freight transportation data of A Railway Group in 2024

序号	日期	需求单号	托运人名称	收货人名称	…	货物品类	货物发送量/t	货运收入/元
1	2024-01-01	***	客户1	***	…	01铁矿石	34	7 144.00
2	2024-01-02	***	客户2	***	…	05工业机械	20	14 121.40
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
164 200	2024-12-30	***	客户86	***	…	07粮食	159	26 361.50

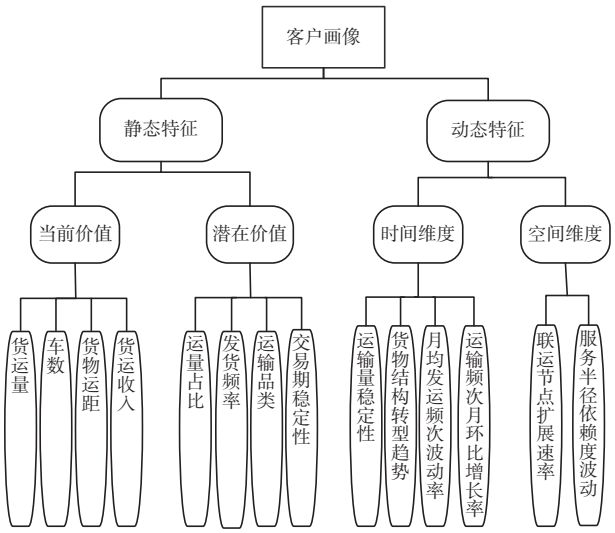

图2　客户画像特征模型

Fig.2　Customer portrait characteristic model

图3所示，得出4个二级指标、8个三级指标的客户特征模型，即潜在价值(运量占比，运输品类)，当前价值(货运量，车数，货运收入)，时间维度(淡季运量，发货频率)，空间维度(货物运距)。

2.2　聚类结果解读与可视化

本研究立足“以客户为中心”，基于铁路物流客户的多维度聚类分析，系统构建客户画像模型。研究选取货运量、货运收入、运量占比、车数、运输品类、淡季运量、发货频率、货物运距等8个关键指标作为画像维度，全面刻画客户行为特征与价值分布。通过提取所有客户发货记录的文本信息，采用词频统计与关键词提取技术，生成客户需求特征图谱，并总结出客户画像特征模型图表[12]，客户核心特征表如表2所示。

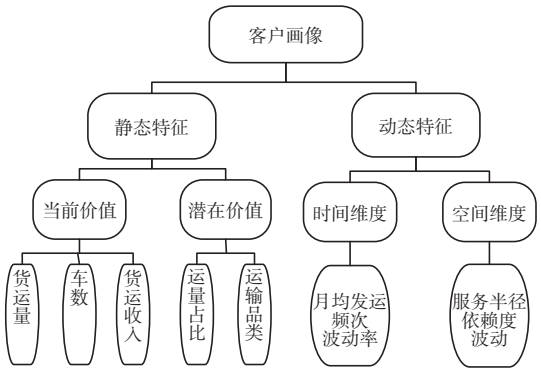

图3　优化后的客户画像特征模型

Fig.3　Optimized customer portrait characteristic model

从货运发送量和货运收入来看，客户群1–战略型客户展现出绝对优势，货运发送量超10 000 t、货运收入超1 000 000元，其规模效应显著，是铁路物流企业稳定的业务支柱；客户群2–利润型客户货运收入同样可观，但发送量处于3 000~10 000 t区间，可能意味着单位货运附加值较高；客户群3–潜力型客户在货运发送量和收入上处于中等水平，具有较大的增长空间；客户群4–零散波动型客户货运量和收入较低，业务规模小且零散。

托运次数与车数方面，战略型客户凭借200次以上的托运次数和300辆以上的车数，彰显出强大的运输需求稳定性和持续性；利润型客户托运次数和车数相对较少，分别为100~200次、100~300辆，运输规模较为适中；潜力型客户托运次数50~100次、车数30~100辆，处于发展阶段；零散波动型客户托运次数和车数极少，分别为0~50次、0~30辆，运输需求零散且不稳定。客户数量上，战略型客户虽仅37个，但贡献的货运量和收入巨大，体现出其单个客户价值极高；利润型客户仅7个，客户数量最少，但个体价值高；潜力型客户有30个，数量相对较多，未来有望成长为更优质客户；零散波动型客户12个，数量不多且单个客户价值有限[13]。

发送货物多样性方面，战略型客户涉及34种

表2　客户核心特征表

Tab.2　Core customer characteristics table

客户类型	客户群1–战略型客户	客户群2–利润型客户	客户群3–潜力型客户	客户群4–零散波动型客户
货运发送量/t	10 000以上	3 000~10 000	1 000~3 000	0~1 000
货运收入/元	1 000 000以上	1 000 000以上	500 000~1000 000	0~500 000
托运次数/次	200以上	100~200	50~100	0~50
车数/辆	300以上	100~300	30~100	0~30
客户数量/个	37	7	30	12
发送货物多样性/种	34	6	5	3
服务半径依赖度波动/%	−5~5	−30~30	−10~20	−50~50
月均发运频次波动率/%	0~5	10~30	0~20	20~50

货物，运输品类丰富，对铁路物流综合运输能力要求高；其余客户群货物种类较少，运输品类相对单一，尤其是零散波动型客户仅涉及3种货物，运输需求较为单一。

服务半径依赖度波动和月均发运频次波动率方面，战略型客户依赖度波动范围最小(−5%~5%)、发运频次波动率在0%~5%，需求稳定且对服务半径依赖性强；利润型客户依赖度波动和发运频次波动率相对较大，业务稳定性稍弱；潜力型客户波动情况居中；零散波动型客户依赖度波动(−50%~50%)和发运频次波动率(20%~50%)最大，业务不稳定，受外部因素影响明显。

2.3　K−means聚类分析

将预处理后的铁路物流客户数据分为静态特征数据集和动态特征数据集；对数据进行标准化处理，以消除不同量对聚类结果的影响。

使用肘部法则，通过计算不同k值下的误差平方和(SSE)，绘制SSE与k的关系图。不同聚类下的误差平方和如图4所示，选取SSE开始急剧下降后趋于平稳的点所对应的k值作为最优聚类数，由图中可以确定最佳聚类数k=4。结合轮廓系数进一步验证所选k值的合理性，确保聚类结果的内聚度和分离度达到最优[14]。

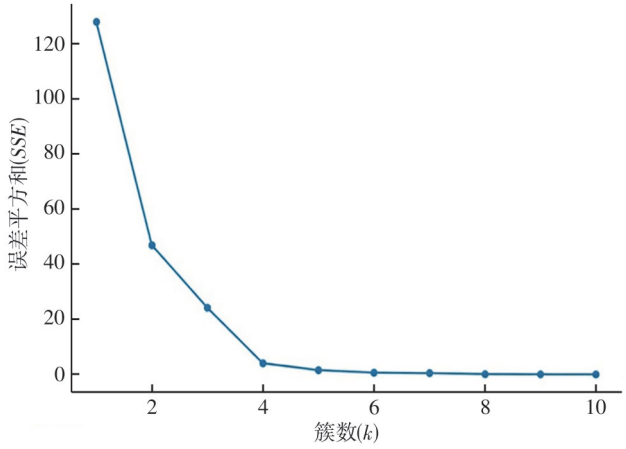

图4　不同聚类下的误差平方和

Fig.4　Sum of squared errors for different clusters

从图4中可以得出，当k=4时SSE曲线出现明显拐点。随着k值继续增大，SSE的下降幅度逐渐趋缓。因此，最终确定最优聚类数为4，该选择在保证较小误差平方和的同时，能够实现对铁路物流客户的精细化分类。在确定聚类数后，采用K−means算法进行聚类分析，铁路物流客户聚类结果如图5所示[15]。

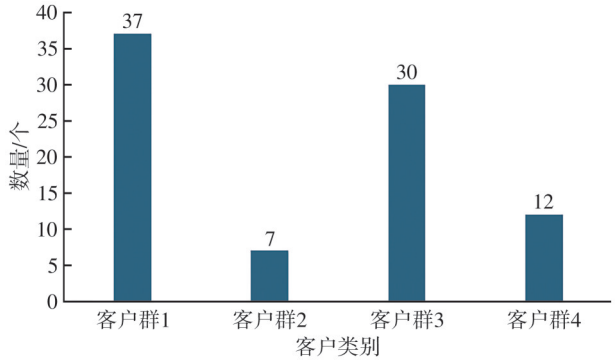

图5　铁路物流客户聚类结果

Fig.5　Clustering results of railway freight customers

为清晰展示细分后不同客户群的差异化特征，本研究采用饼状图和Tagul词云工具进行可视化呈现。各客户群的核心特征分析如下[16]。

2.3.1　客户群1核心客户

客户群1(占比43.02%，37家客户)是铁路货运的核心支柱，具有规模大、稳定性强的显著特征，客户群1画像图如图6所示。该群体以金属矿石运输为主(占比89.57%，年运量超297万t)，客户群1中不同品类货物占比如图7所示，单客户年均货运量超1万t，收入贡献均超100万元。其运输需求高度稳定(发运频次波动仅±3%)，客户群1月均发运频次波动率如图8所示，服务半径集中(依赖度波动±5%以内)，客户群1服务半径依赖度如图9所示，供应链网络抗风险能力强。建议将其定位为“战略核心客户”，提供专属客户经理、长期协议和优先运力保障等VIP服务，以巩固这一高价值客户群的长期合作[6]。



图6　客户群1画像图

Fig.6　Portrait of customer group 1

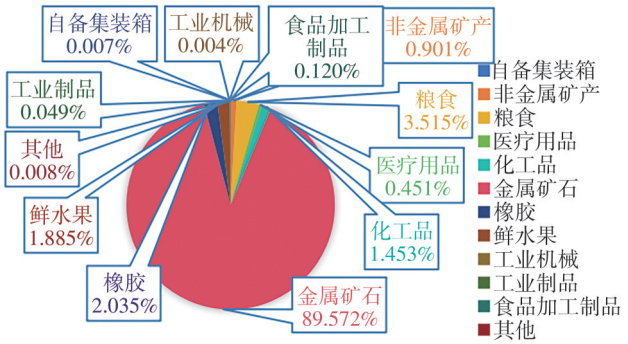

图7　客户群1中不同品类货物占比

Fig.7　Proportion of different cargo categories in customer group 1

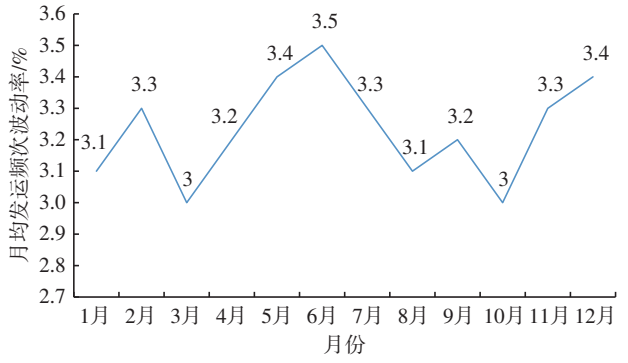

图8　客户群1月均发运频次波动率

Fig.8　Monthly shipment frequency volatility rate for customer group 1

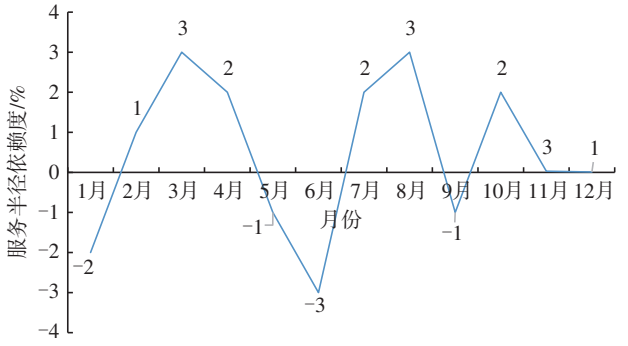

图9　客户群1服务半径依赖度

Fig.9　Service radius dependency of customer group 1



图10　客户群2画像图

Fig.10　Portrait of customer group 2

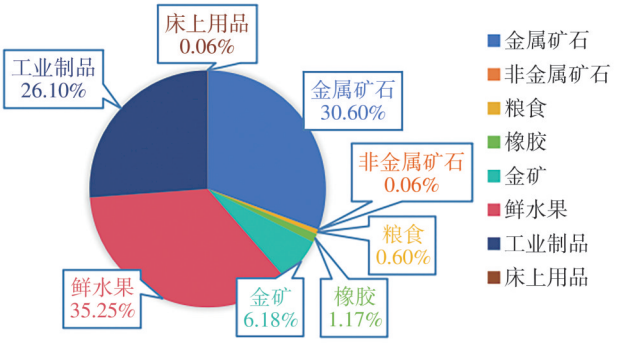

图11　客户群2中不同品类货物占比

Fig.11　Proportion of different cargo categories in customer group 2

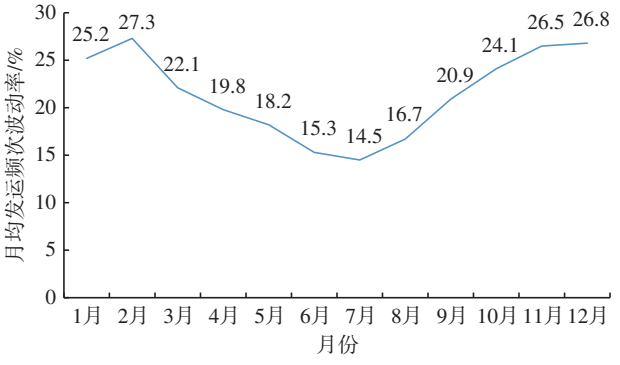

图12　客户群2月均发运频次波动率

Fig.12　Monthly shipment frequency volatility rate for customer group 2

2.3.2　客户群2重要客户

客户群2(占比8.13%，8家客户，客户群2画像图如图10所示)具有货源多样(鲜水果占35.25%，客户群2中不同品类货物占比如图11所示)、需求波动大(发运频次波动±30%，客户群2月均发运频次波动率如图12所示)和区域依赖性强但不稳定(核心区域占比60%~70%，依赖度波动±25%，客户群2服务半径依赖度如图13所示)的特点。虽然单客户年收入超100万元，但其高波动性增加了运力调配难度。建议将其定位为“重要利润型客户”，通过弹性运力、动态定价和重点品类时效保障等措施，在确保收益的同时应对运营挑战。

2.3.3　客户群3潜力客户

客户群3(占比34.89%，30家客户，客户群3画像图如图14所示)是铁路货运的中坚力量，以矿建材料为主(占比80.98%，客户群3中不同品类货物占比如图15所示)，单客户年均货运量1 000~3 000 t，收入贡献50~100万元。该群体需求波动适中(±20%，客户群3月均发运频次波动如图16所示)，服务半

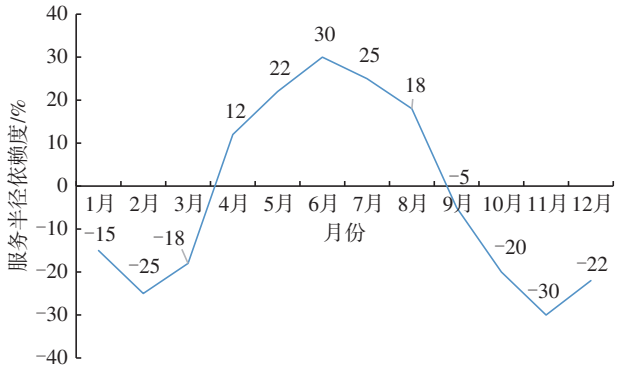

图13　客户群2服务半径依赖度

Fig.13　Service radius dependency of customer group 2



图14　客户群3画像图

Fig.14　Portrait of customer group 3

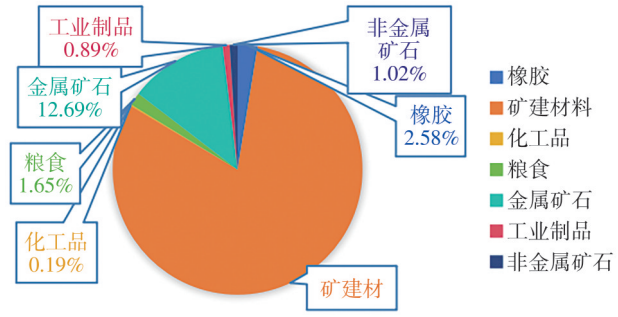

图15　客户群3中不同品类货物占比

Fig.15　Proportion of different cargo categories in customer group 3

径稳定(依赖度波动-10%~19%，客户群3服务半径依赖度如图17所示)，展现出良好的发展潜力。建议将其定位为“潜力型客户”，通过性价比服务方案、品类需求拓展和年度框架协议等措施，培育其成为未来稳定的重要收入来源。

2.3.4　客户群4普通客户

客户群4(占比13.96%，12家客户，客户群4画像图如图18所示)是典型的零散波动型客户，以重晶石和医疗用品运输为主，单客户年均货运量不足1 000 t，收入贡献低于50万元，客户群4中不同品

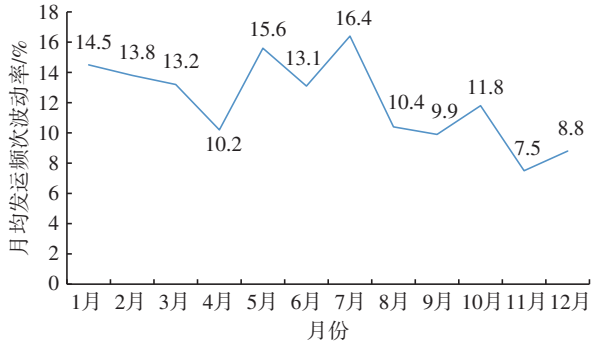

图16　客户群3月均发运频次波动

Fig.16　Monthly shipment frequency volatility for customer group 3

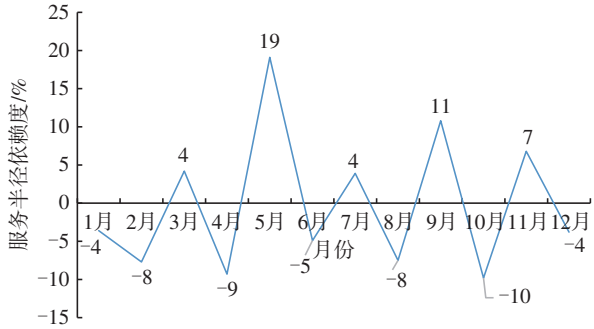

图17　客户群3服务半径依赖度

Fig.17　Service radius dependency of customer group 3



图18　客户群4画像图

Fig.18　Portrait of customer group 4

类货物占比如图19所示。该群体具有极高的需求波动性(±50%，客户群4月均发运频次波动如图20所示)和服务半径不稳定性(波动-47%~52%，客户群4服务半径依赖度如图21所示)。建议将其定位为“普通波动型客户”，采取标准化服务流程和灵活运力调配策略，重点通过线上平台简化服务流程，并针对特定品类开发标准化的小批量运输产品，以低成本方式满足其零散需求。

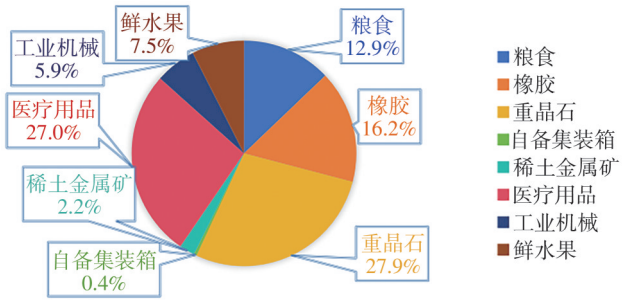

图19　客户群4中不同品类货物占比

Fig.19　Proportion of different cargo categories in customer group 4

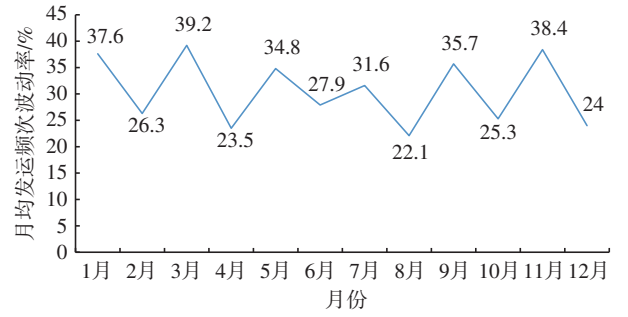

图20　客户群4月均发运频次波动

Fig.20　Monthly shipment frequency volatility for customer group 4

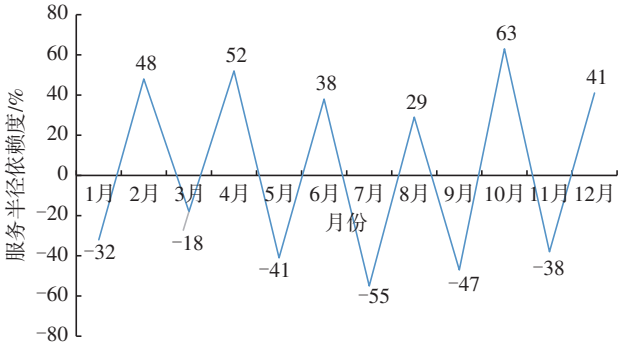

图21　客户群4服务半径依赖度

Fig.21　Service radius dependency of customer group 4

3　结束语

本研究通过构建铁路物流客户画像的动静态特征融合体系，突破传统单维建模局限，完成了静态特征与动态特征的结合分析。通过具体的铁路物流客户静态属性和行为特征进行精细指标划分，采用聚类算法优化与验证，使用肘部法则、轮廓系数确定最优聚类数k=4，结合误差平方和准则优化初始聚类中心，显著提升了K-means算法的稳定性和聚类精度。多种指数及评估指标表明，优化后的算法在客户分类的内聚度和分离度上表现更优、更能解决特征间的差异性。通过案例分析进行客户群体特征识别，基于聚类结果，将铁路货运客户划分为战略型客户、利润型客户、潜力型客户和零散波动型客户4类。各类客户在货运量、收入贡献、运输需求波动性及货物结构等方面呈现显著差异，为差异化服务策略制定提供了明确依据。本研究在未来可以进一步扩充数据样本量，进一步优化算法以提升分析准确性与合理性。

参考文献：

[1] 刘　浩，申嘉琪，张　戎．“公转铁”背景下铁路货运市场化定价策略[J]．铁道货运，2022，40(10)：54-60.
LIU Hao，SHEN Jiaqi，ZHANG Rong. Market-Oriented Pricing Strategy of Railway Freight under the Background of Modal Shift from Road to Rail[J]. Railway Freight Transport，2022，40(10)：54-60.

[2] 秦　胜．基于客户需求特征的铁路运输生产计划创新研究[J]．铁道运输与经济，2017，39(10)：42-47.
QIN Sheng. Study on the Railway Transportation Production Plan innovation based on Customer Demand Characteristics[J]. Railway Transport and Economy，2017，39(10)：42-47.

[3] 李昊明．基于多维特征的载客区域推荐[D]．武汉：武汉大学，2022.

[4] 王　菲．基于客户特征机器学习的铁路货运差异化服务匹配方法研究[J]．铁道运输与经济，2025，47(5)：186-194.
WANG Fei. Differentiated Service Matching Method of Railway Freight Based on Customer Characteristic-Oriented Machine Learning[J]. Railway Transport and Economy，2025，47(5)：186-194.

[5] 赵芳璨．基于深度学习的铁路短期货运量预测及客户分析研究[D]．北京：北京交通大学，2023.

[6] 贾倩新．铁路货运客户综合评价及资源优化配置[D]．兰州：兰州交通大学，2023.

[7] 龚德鹏．基于数据挖掘的铁路货运潜在客户识别研究[D]．成都：西南交通大学，2016.

[8] 张启明，陆建华，徐建栋，等．基于朴素贝叶斯算法的客户画像分析[J]．无线互联科技，2020，17(1)：122-125.

[9] 杨　翌．铁路货运客户价值预测与潜在客户识别方法研究[D]．北京：北京交通大学，2023.

[10] 胡双全．基于大数据的铁路货运大客户画像设计与应用研究[D]．天津：天津商业大学，2021.

[11] 赵惜茹．基于大数据技术的银行客户画像构建的研究与应

（下转第204页）