

DOI:10.11918/202211034

# 危险货物道路运输个性化路径推荐方法

方琼<sup>1</sup>, 钱大琳<sup>1</sup>, 陈心如<sup>2</sup>, 李思贤<sup>1</sup>

(1. 综合交通运输大数据应用技术交通运输行业重点实验室(北京交通大学), 北京 100044;  
2. 武汉地铁运营有限公司, 武汉 430035)

**摘要:** 为加强危险货物道路运输风险源头管控, 以危险运输车辆行驶轨迹数据为分析对象, 研究安全、经济且符合企业自身偏好的道路运输路径优化选择问题, 提出了基于偏好、上下文感知的危险货物道路运输个性化路径推荐方法。首先对危险运输车辆历史轨迹数据进行处理, 通过提取路径安全性和经济性特征学习危险运输企业的路径偏好, 在此基础上, 综合考虑偏好向量间的距离和方向相似性, 提出了改进的  $K$ -means 偏好聚类算法(improved  $K$ -means clustering algorithm based on distance and direction similarity measurement, DDM- $K$ -means), 获取了路径偏好类别; 其次, 依据运输任务执行的时间、天气、运距三方面信息, 建立了路径上下文向量, 并运用 Rock 聚类算法划分路径的上下文类别, 与路径偏好类别共同构成路径信息; 最终, 基于神经协同过滤提出了危险货物道路运输路径选择优化算法(optimal route selection algorithm based on neural collaborative filtering, NCF-ORS), 得到了危险运输企业对各路径类别的偏好排序, 从而为企业推荐最优路径。与基线算法比较分析, 结果表明危险货物道路运输个性化路径推荐方法 < DDM- $K$ -means, NCF-ORS >, 平均绝对百分比误差最低。研究结果有助于挖掘车辆轨迹数据中更多的潜在信息, 提升个性化路径推荐能力, 可为危险运输企业的选线问题提供决策支持。

**关键词:** 危险品运输; 路径推荐; 神经协同过滤; 偏好聚类算法; 上下文感知

中图分类号: U492.336.3 文献标志码: A 文章编号: 0367-6234(2024)07-0055-08

## Personalized route recommendation method for road transport of hazardous materials

FANG Qiong<sup>1</sup>, QIAN Dalin<sup>1</sup>, CHEN Xinru<sup>2</sup>, LI Sixian<sup>1</sup>

(1. Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport (Beijing Jiaotong University), Beijing 100044, China; 2. Wuhan Metro Operation Co., Ltd., Wuhan 430035, China)

**Abstract:** To strengthen the management and control of the source of risk for road transport of hazardous materials, this paper takes the trajectory data of hazardous materials transport vehicles as the analysis object, and studies the problem of optimal selection of road transport routes which is safe, economical and in line with the preference of enterprises, a context-aware, preference-based personalized route recommendation method for road transport of hazardous materials is proposed. Firstly, the historical trajectory data of hazardous materials transport vehicles is processed, and the route preferences of enterprises are learned by extracting route safe and economical features. On this basis, considering the distance and direction similarity between preference vectors, an improved  $K$ -means clustering algorithm (DDM- $K$ -means) is proposed to obtain the categories of route preference. Secondly, according to the time, weather, and distance of the transportation tasks, the route context vectors are established. Rock clustering algorithm is used to classify the categories of route context, combined with the categories of route preference to form the categories of route. Finally, based on neural collaborative filtering, an optimal route selection algorithm (NCF-ORS) is proposed, and the preference ranking of hazardous materials road transport enterprises for route categories is obtained to recommend the optimal route for enterprises. Comparing our method with the baseline algorithms, the results showed that the personalized route recommendation method < DDM- $K$ -means, NCF-ORS > proposed in this paper had a lower mean absolute percentage error. Therefore, the research in this paper is helpful to mine more potential information from vehicle's trajectory data, with stronger personalized route recommendation capabilities, and can provide decision support for route selection of hazardous materials road transport enterprises.

**Keywords:** transportation of hazardous materials; route recommendation; neural collaborative filtering; preference clustering algorithm; context-aware

收稿日期: 2022-11-09; 录用日期: 2022-12-24; 网络首发日期: 2024-06-13

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20240612.1632.004>

基金项目: 国家自然科学基金(62272030, 52072289)

作者简介: 方琼(2000—), 女, 硕士研究生; 钱大琳(1963—), 女, 教授, 博士生导师

通信作者: 钱大琳, dlqian@bjtu.edu.cn

危险货物是指具有易燃、易爆、毒害、腐蚀或放射性等性质的物品,一旦发生事故会造成灾难性的损失,严重威胁着沿途居民的生命安全和周边的生态环境,是重大的社会公共安全问题之一。因此,科学地优化危险货物道路运输线路,确保危险货物运输安全、快速和经济,对于危险货物运输风险管理具有重要的理论研究和实际意义<sup>[1]</sup>。

既有的基于图搜索的人工路径规划研究<sup>[2]</sup>和智能仿生路径规划研究<sup>[3]</sup>可为用户推荐客观的优化路径,但缺乏对规划者偏好的考虑。卫星定位系统在危险货物道路运输行业的普及以及推荐算法的提出,使得运用车载终端设备采集的车辆轨迹数据深入分析危货道路运输企业的路径选择偏好成为可能。由此,众多学者开展了个性化路径规划研究<sup>[4-8]</sup>。文献[6]将基于内容的推荐和协同过滤算法相结合,构建了一个混合推荐算法,但没有考虑运输时间、天气等运输任务相关的信息对路径决策的影响。文献[7]进一步完善了路径推荐方法,将运输任务相关的信息作为路径上下文,纳入路径推荐方法中,但提出的路径推荐算法复杂度高,不适用于大规模的学习。文献[8]利用基于矩阵分解的协同过滤技术,预测驾驶员对路径的偏好,实现了上下文感知的个性化路径推荐,但矩阵分解难以挖掘用户和项目间的非线性特征,在学习两者潜在交互时存在局限性<sup>[9]</sup>。

此外,通过聚类获取路径的偏好类别是实现路径推荐的前提。目前的众多聚类算法中  $K$ -means 算法应用最广,并衍生出了大量不同的改进算法<sup>[10]</sup>。其中,改变向量间的相似性度量,如文献[11]用马氏距离代替欧氏距离,可以不受特征量纲不同的影响,避免特征属性之间的相关性对距离度量产生干扰。文献[12]采用余弦相似度作为聚类算法的度量标准,可以区分不同个体间方向上的差异,在不同适用场景下均能有效提升算法的聚类效果。但上述方法仅单一地对向量间的距离或方向相似性进行度量,不适用于基于偏好的路径聚类。

鉴于此,本文提出了一种基于偏好、上下文感知的危货道路运输个性化路径推荐方法,为危货道路运输企业提供规范的路径决策服务。研究中同时考虑偏好向量间距离和方向相似性,提出了一种改进的  $K$ -means 聚类算法 DDM- $K$ -means,用于划分路径的偏好类别;并提出了基于神经协同过滤的路径选择优化算法 NCF-ORS,实现了对危货道路运输企业

与路径间复杂交互的线性与非线性建模,可以挖掘企业选择运输路径的潜在偏好。

## 1 问题描述与技术路线

### 1.1 问题描述

本文研究的是针对特定的危货道路运输任务,为企业推荐安全、经济且符合企业自身偏好的道路运输最优路径问题,即:结合该项运输任务的时间、天气和距离等特征信息,以及危货道路运输企业路径选择的偏好,根据从危货运输车辆历史轨迹数据中建立的候选路径集,确定执行该项任务的安全、经济优化路径。相关符号定义如下:

1) 令  $G = (V, E)$  为道路网络,其中  $V$  和  $E$  分别是顶点集和边集。 $v_i \in V$  表示顶点, $e_j \in E$  表示有向路段。

2) 令  $R_{\text{all}} = \{R_1, R_2, \dots, R_i, \dots, R_{n_r}\}$  为历史路径集,其中  $R_i$  表示第  $i$  条路径; $n_r$  表示历史路径的总条数, $1 \leq i \leq n_r$ 。

3) 将危货道路运输企业视为推荐算法中的用户  $u$ ;路径类别视为项目  $i$ ,用  $i = \langle m, n \rangle$  表示,其中  $m$  为危货道路运输企业选择路径的偏好类别; $n$  为表示时间、天气和距离等危货运输任务特征的路径上下文类别。用户集用  $U$  表示,项目集用  $I$  表示。

4) 将用户  $u$  是否访问过项目  $i$  中的路径,视为用户与该项目是否存在交互行为,用  $y_{ui}$  表示交互次数。需要注意的是, $y_{ui} = 0$  表示用户  $u$  没有访问过项目  $i$ ,但并不代表用户  $u$  不喜欢项目  $i$ 。定义用户-项目交互矩阵为  $\mathbf{Y} = \{y_{ui} | u \in U, i \in I\} \in R^{E \times F}$ ,其中  $E$  和  $F$  分别表示用户和项目的数量,且  $F = M \times N$ , $M$  为路径偏好类别数, $N$  为路径上下文类别数。

### 1.2 技术路线

本文所提出的危货道路运输个性化路径推荐方法,具体包括:危货道路运输企业路径偏好特征提取和基于偏好的路径聚类;路径上下文识别及其聚类;企业-路径的交互关系挖掘和基于偏好的个性化路径推荐。其中危货道路运输企业路径偏好指运输路径的安全性和经济性特征;上下文指运输任务执行的时间、天气、运距信息。最终通过与基线算法的对比分析验证该方法的有效性,技术路线如图 1 所示。图中路径偏好聚类和路径上下文聚类均采用主成分分析实现聚类效果可视化,PC1 和 PC2 分别表示从数据差异性最大和次大的方向提取出来的主成分。

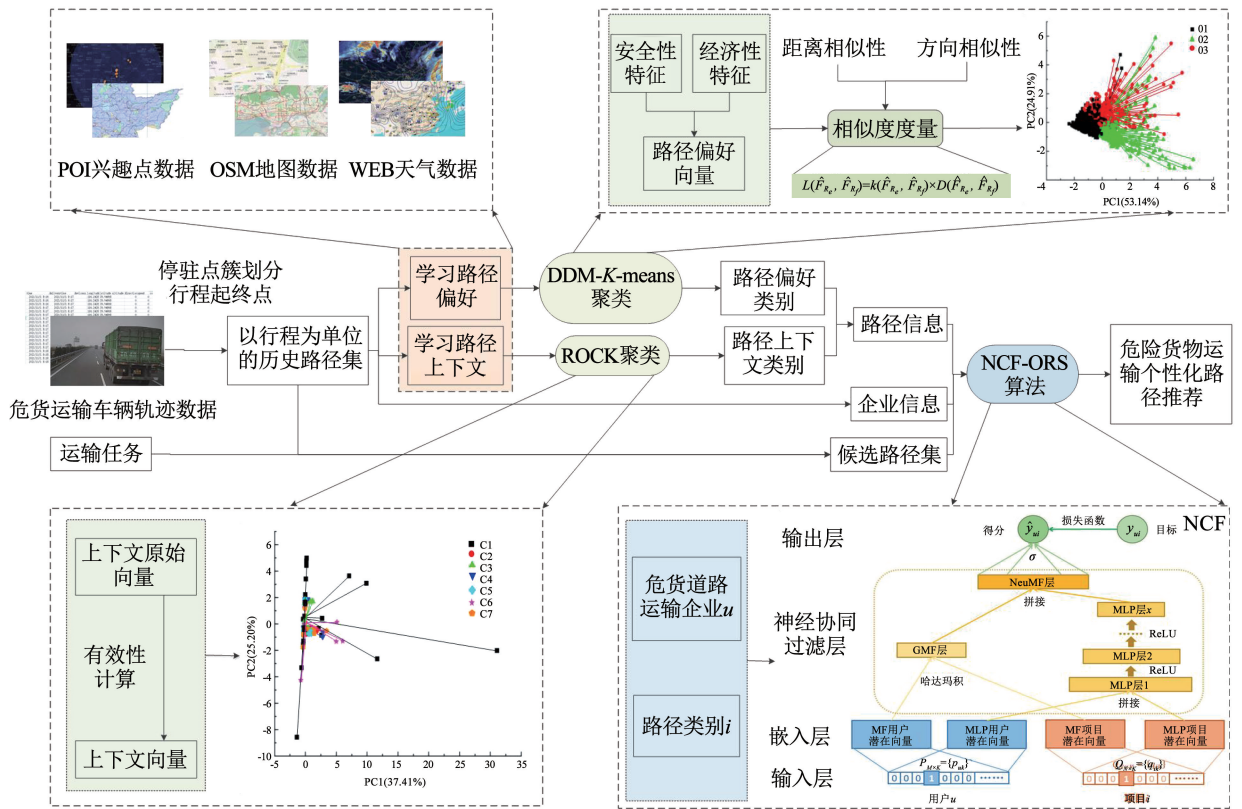


图1 技术路线

Fig.1 Methodology route

## 2 基于偏好、上下文感知的路径聚类

实现个性化路径推荐的基础是获取企业-路径的交互信息,企业可用企业代号表示,而路径类别无法直接从轨迹数据中得到。因此,本文需要进行基于偏好、上下文感知的路径聚类,得到历史路径集中所有路径的偏好类别和上下文类别,共同构成路径类别。

### 2.1 路径偏好聚类

危险货物具有易燃易爆等化学特性,运输过程中存在特殊风险,因此危货道路运输企业在追求自身经济利益的同时也必须将运输安全放在首位,在进行路径规划时需要将运输成本和运输安全进行平衡决策,最终的选线结果也反映了不同运输企业的路径偏好。为对其进行定量描述,本文从以下两个方面提取路径特征:1)安全性特征。低等级路段占比指低等级路段长度在整个运输路径长度中的占比,用 $\hat{f}^{(1)}$ 表示;单转向次数指每公里危货运输车辆的平均转向次数,用 $\hat{f}^{(2)}$ 表示。2)经济性特征。距离绕行度指路径的实际运输长度相对于起讫点间的最短运输距离的绕行值,用 $\hat{f}^{(3)}$ 表示;耗时比指路径的实际运输时长相对于起讫点间的最短运输时间的比值,用 $\hat{f}^{(4)}$ 表示。

根据上述4个指标为历史路径集  $R_{all}$  中的每条

路径提取特征信息,并构建偏好向量。将路径  $R_i$  的偏好向量表示为  $\hat{F}_{R_i} = (\hat{f}_{R_i}^{(1)}, \hat{f}_{R_i}^{(2)}, \hat{f}_{R_i}^{(3)}, \hat{f}_{R_i}^{(4)})$ , 则所有历史运输路径的偏好向量集合为  $\hat{F} = \{\hat{F}_{R_i} | R_i \in R_{all}\}$ 。

#### 2.1.1 偏好相似性度量

向量间的相似性度量对聚类效果有显著影响。欧氏距离是最常见的相似性度量,并在此基础上衍生出了加权欧氏距离、马氏距离<sup>[11]</sup>等,可用于度量两个体间数值上的差异,即距离相似性;余弦相似度<sup>[12]</sup>是指高维空间内两向量夹角的余弦值,可用于度量两个体间方向上的差异。但既有的单一的相似性度量无法同时考虑危货道路运输企业在选线时对各路径特征的偏好倾向(方向相似性)和偏好程度(距离相似性)。因此,本文使用余弦相似度衡量偏好向量间的方向相似性,马氏距离衡量偏好向量间的距离相似性,构建了综合考虑距离和方向相似性的度量模型,具体步骤如下:

1)方向相似性度量。本文将路径偏好向量中各特征值与平均值之差作为每条路径的倾向性向量,计算倾向性向量间的余弦相似度,其正负体现了路径偏好向量间方向上的相似性,并引入同倾向性权重进行表示,计算公式为

$$\cos(\hat{F}_{R_c}', \hat{F}_{R_t}') = \sum_{i=1}^4 [(\hat{f}_{R_c}^{(i)} - \bar{f}^{(i)}) \times (\hat{f}_{R_t}^{(i)} - \bar{f}^{(i)})] /$$

$$\left[ \sqrt{\sum_{t=1}^4 (\hat{f}_{R_e}^{(t)} - \bar{f}^{(t)})^2} \times \sqrt{\sum_{t=1}^4 (\hat{f}_{R_f}^{(t)} - \bar{f}^{(t)})^2} \right] \quad (1)$$

$$k(\hat{F}_{R_e}, \hat{F}_{R_f}) = \begin{cases} 1, & \cos(\hat{F}_{R_e}, \hat{F}_{R_f}) > 0 \\ 2, & \cos(\hat{F}_{R_e}, \hat{F}_{R_f}) \leq 0 \end{cases} \quad (2)$$

式中： $\hat{F}_{R_i}'$ 为路径  $R_i$  的倾向性向量， $\cos(\hat{F}_{R_e}, \hat{F}_{R_f})$ 为路径  $R_e$  与路径  $R_f$  倾向性向量间的余弦相似度， $\hat{f}_{R_i}^{(t)}$ 为路径  $R_i$  的偏好向量中第  $t$  个特征的值， $\bar{f}^{(t)}$ 为所有历史路径的偏好向量中第  $t$  个特征的平均值， $\hat{F}_{R_e}$ 、 $\hat{F}_{R_f}$ 分别为路径  $R_e$ 、 $R_f$  的偏好向量， $k(\hat{F}_{R_e}, \hat{F}_{R_f})$ 为偏好向量  $\hat{F}_{R_e}$  与  $\hat{F}_{R_f}$  之间的同倾向性权重。通过计算可以得到每两个路径偏好向量间的同倾向性权重。

2) 距离相似性度量。马氏距离不受路径特征量纲不同的影响，能有效计算路径偏好向量之间的距离，具体计算公式为

$$D(\hat{F}_{R_e}, \hat{F}_{R_f}) = \sqrt{(\hat{F}_{R_e} - \hat{F}_{R_f})^T \Sigma^{-1} (\hat{F}_{R_e} - \hat{F}_{R_f})} \quad (3)$$

式中： $D(\hat{F}_{R_e}, \hat{F}_{R_f})$ 为偏好向量  $\hat{F}_{R_e}$  与  $\hat{F}_{R_f}$  之间的马氏距离， $\Sigma$ 表示协方差矩阵。通过上述计算可以得到每两个路径偏好向量之间的马氏距离。

3) 偏好相似性度量。使用同倾向性权重加权马氏距离对偏好向量间的相似性进行度量，用  $L(\hat{F}_{R_e}, \hat{F}_{R_f})$ 表示，计算公式为

$$L(\hat{F}_{R_e}, \hat{F}_{R_f}) = k(\hat{F}_{R_e}, \hat{F}_{R_f}) \times D(\hat{F}_{R_e}, \hat{F}_{R_f}) \quad (4)$$

### 2.1.2 DDM-K-means 聚类算法

根据上述相似性度量模型，本文提出一种改进的  $K$ -means 聚类算法 (improved  $K$ -means clustering algorithm based on distance and direction similarity measurement, DDM- $K$ -means)。该算法借鉴  $K$ -means++ 算法选择初始类簇的方法，并采用 2.1.1 节中所提出的相似性度量模型代替欧氏距离，进一步提高了聚类结果的准确性和可解释性。本文使用 DDM- $K$ -means 算法对路径偏好向量进行聚类，算法的实现步骤如下：

1) 设定路径偏好类别数为  $M$  (即聚类的类簇数)，随机选取一个路径偏好向量作为第 1 个初始类簇中心。

2) 找到离第一个初始类簇中心最远的向量，标记为第 2 个初始类簇中心。并按照最远原则依次选取，最终得到  $M$  个初始类簇中心。

3) 使用上述相似性度量模型 (同倾向性权重加权马氏距离) 计算其余偏好向量到每一个类簇中心的距离，如式 (1) ~ (4) 所示。

4) 将路径偏好向量划分到距离最近的类簇中。

5) 将同一类簇中的所有向量求平均，得到新的类簇中心。

6) 返回步骤 4) 迭代，直至各类簇中心不再发生改变，迭代停止。

由此可将路径偏好向量划分为  $M$  个类别，路径  $R_i$  的路径偏好类别可以表示为  $m_{R_i}$ 。遍历历史路径数据集，得到所有路径的路径偏好类别。

## 2.2 路径上下文聚类

运输任务特征会影响危货道路运输企业的选线偏好。例如，恶劣天气下企业更倾向选择高速公路等。因此本文采用时间、天气、运距等运输任务特征信息作为上下文，考虑了出发时刻、出发日期、天气状况、运输距离等因素，分别用向量  $S^{(1)}$ 、 $S^{(2)}$ 、 $S^{(3)}$ 、 $S^{(4)}$  表示，则运输路径  $R_i$  的上下文信息可以表示为上下文原始向量  $S_{R_i} = (S_{R_i}^{(1)}, S_{R_i}^{(2)}, S_{R_i}^{(3)}, S_{R_i}^{(4)})$ 。但上下文中还隐含了部分无效信息，例如对于某个企业，节假日对选线没有影响，但恶劣天气下却选择了不同路径。因此本文使用基尼指数计算上下文有效性，对上下文向量进行修正。有效性加权后的上下文向量表示为  $\hat{S}_{R_i} = (\hat{S}_{R_i}^{(1)}, \hat{S}_{R_i}^{(2)}, \hat{S}_{R_i}^{(3)}, \hat{S}_{R_i}^{(4)})$ ，上下文向量集合为  $\hat{S} = \{\hat{S}_{R_i} | R_i \in R_{\text{all}}\}$ 。

本文中的上下文信息均为类别型数据，如是否高峰时段的上下文为  $S^{(1)} = \{0, 1\}$ ，0 表示出发时刻处于非高峰时段，1 则表示出发时刻处于高峰时段。Rock 算法属于凝聚型的层次聚类，是一种鲁棒的适用于类别型数据的经典聚类算法，因此本文使用 Rock 算法实现上下文聚类。需要注意的是，若只为历史运输路径的上下文向量聚类，无法直接判断不在集合  $\hat{S}$  中的全新上下文向量的类别，因此需要列举出所有可能出现的上下文向量，创建集合  $\hat{S}_{\text{all}}$ ，则  $\hat{S} \subseteq \hat{S}_{\text{all}}$ ， $\hat{S}_{\text{all}} = \{\hat{S}_1, \hat{S}_2, \dots, \hat{S}_x\}$ 。获取  $\hat{S}_{\text{all}}$  中每一个上下文向量对应的上下文类别，并保存为上下文类别表。定义上下文类别为  $N$ ，根据历史运输路径  $R_i$  对应的上下文向量  $\hat{S}_{R_i}$  在上下文类别表中查找其上下文类别，用  $n_{R_i}$  表示。遍历历史路径数据集，得到所有路径的上下文类别，与路径偏好类别一起共同构成路径类别。

## 3 N-ORS 路径选择优化算法

在路径规划方面，国内外已开展了大量的研究工作。Dijkstra、A\*、Floyd 算法等基于图搜索的人工路径规划方法<sup>[2]</sup>通常提供的是起讫点间的最短路径；遗传算法、蚁群算法等智能仿生路径规划方法<sup>[3]</sup>可模仿生物群体行为或生态机制为用户推荐

较优路径,但均忽略了路径规划者的个体偏好。推荐算法能够发现用户或项目间的相关性,然后基于以往的用户-项目交互记录,为用户推荐符合其偏好的项目,近年来也被逐渐应用于路径规划领域。其中,基于矩阵分解的协同过滤技术应用最为广泛,但其难以挖掘危货道路运输企业与路径之间的非线性关联。

神经协同过滤(NCF)<sup>[9]</sup>包括输入层、Embedding层、神经协同过滤层和输出层。该模型是对矩阵分

解的改进,将广义矩阵分解(generalized matrix factorization, GMF)和多层感知机(multi-layer perceptron, MLP)融合,能够实现对用户-项目的线性和非线性建模。由此,本文提出了一种基于神经协同过滤的危险货物道路运输路径选择优化算法(optimal route selection algorithm based on neural collaborative filtering, NCF-ORS),用于实现个性化路径推荐,引入的路径上下文附加信息也有效缓解了数据稀疏性问题,算法框架如图2所示。

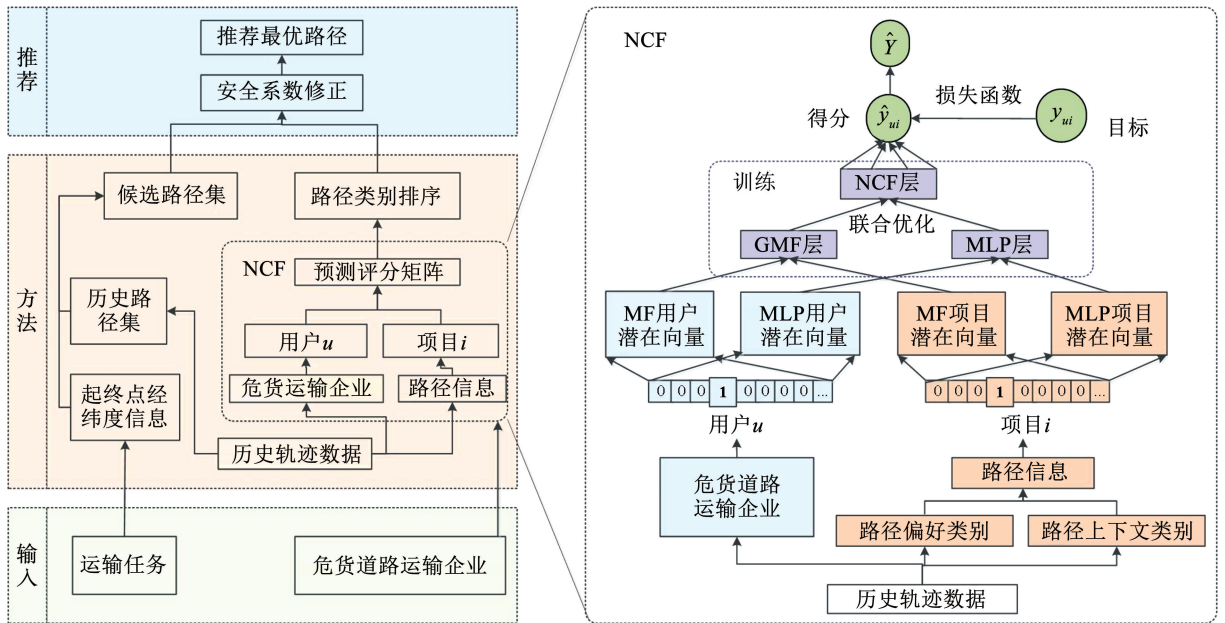


图2 NCF-ORS 算法框架

Fig. 2 NCF-ORS algorithm framework

### 3.1 NCF 模型训练

本文根据历史轨迹数据挖掘到的路径信息,训练 NCF 模型以获取危货道路运输企业对路径的预测评分集合。训练好的模型用于 NCF-ORS 算法,可为输入的企业得到其对各路径类别的偏好排序,以便从候选路径集中选择最优路径。NCF 模型训练具体步骤如下:

1) 模型输入。本文 2.1 节和 2.2 节中分别得到了路径偏好类别和路径上下文类别,将其构成路径类别(项目  $i$ ),与危货道路运输企业(用户  $u$ )一起输入到模型中。

2) 模型训练。将访问路径  $R_i$  的企业用  $u_{R_i}$  表示,为路径  $R_i$  生成信息集  $C_{R_i} = \{u_{R_i}, m_{R_i}, n_{R_i}\}$ ,则所有路径的信息集合可表示为  $C = \{C_{R_i} | R_i \in R_{all}\}$ 。进而得到了用户  $u$  与项目  $i$  之间的交互次数  $y_{ui}$ ,构建用户-项目交互矩阵  $Y \in R^{E \times F}$ ,将其作为标签对模型进行训练,选择交叉熵损失函数来学习模型参数。

3) 模型输出。通过训练 NCF 模型,得到了用户对项目的预测评分集合  $\hat{Y}$ ,可以表示为  $\hat{Y} = \{\hat{y}_{ui} | u \in$

$U, i \in I\}$ 。其中,需要判断企业  $u$  是否为冷启动用户,遍历企业集合  $U_{all} = \{U_1, U_2, \dots, U_k, \dots, U_K\}$ ,若该企业在企业集合中,即  $u \in U_{all}$ ,则可直接从预测评分矩阵  $\hat{Y}$  中提取  $\hat{y}_{ui}$ ;否则,企业为冷启动用户,通过其他企业来计算该企业的评分值,令  $\hat{y}_{ui} = (1/K) \sum_{k=1}^K \hat{y}_{|U_k, i}$ 。

### 3.2 N-ORS 算法框架

**输入:**将企业代号与起讫点、时间、天气和距离等运输任务信息作为算法的输入。

**方法:**运用危货道路运输车辆轨迹数据建立历史路径集,根据运输任务起讫点经纬度信息从历史路径集中获取起讫点间的所有历史路径,并构造为候选路径集  $R_{can} = \{R_1, R_2, \dots, R_l, \dots, R_L\}$ ;运用 NCF 模型从历史轨迹数据中挖掘企业-路径的交互关系,得到该危货道路运输企业对各路径类别的预测评分集合  $\hat{Y}$ ;将  $\hat{Y}$  中的评分依次递减,获得危货道路运输企业在特定上下文下对各路径偏好类别的偏好排序,记为  $k_i$ ,其中  $i \in I$ 。

**推荐:**根据候选路径集  $R_{\text{can}}$  与危货道路运输企业对各路径类别的偏好排序  $k_i$ , 完成安全、经济且符合企业自身偏好的道路运输路径优化选择。

将候选路径集  $R_{\text{can}}$  中路径  $R_l$  的类别记为  $i_{R_l}$ 。若  $i_{R_l} = i$ , 则  $k_{R_l} = k_i$ 。基于危险货物的特殊性, 推荐路径时应倾向于推荐安全系数较高的路径, 因此本文利用路径的安全性特征指标对  $k_{R_l}$  进行修正, 计算公式为

$$k'_{R_l} = (\hat{f}_{R_l}^{(1)} + \hat{f}_{R_l}^{(2)}) \times k_{R_l} \quad (5)$$

式中:  $\hat{f}_{R_l}^{(1)}$  为路径  $R_l$  的低等级路段占比,  $\hat{f}_{R_l}^{(2)}$  为路径  $R_l$  的单位转向次数,  $k'_{R_l}$  为路径  $R_l$  的最终路径排序。根据修正后的路径排序  $k'_{R_l}$  将候选路径集中的路径从低到高排列, 并选择路径排序值最小的路径  $R_{\text{final}} = \{R_l | \min k'_{R_l}\}$  推荐给企业。

## 4 实验结果与分析

### 4.1 数据集

本文数据来源于 23 家危货道路运输企业在 2021 年 11 月份的轨迹数据, 原始文件是以危货道路运输车辆代号命名的轨迹序列, 每个文件内有一个月的连续轨迹点信息。轨迹数据采集间隔为 5 s, 包括时间、经纬度、速度和累积行驶里程等。

通过对原始文件进行筛选, 发现原始轨迹数据中不存在字段缺失和明显异常点等情况, 因此本文只需通过 Haversine 公式计算距离去除漂移点来完成数据清洗。要得到以行程为单位的路径集, 首先需要判断行程的起讫点。本文通过对轨迹数据进行聚类得到停驻点簇, 并将停留时间超过 15 min 的停驻点簇视为起讫点。为解决实际行程起讫点经纬度与停驻点簇结果存在偏差的问题, 本文扩大了停驻点簇的坐标范围(经纬度  $\pm 0.003$ ), 将落入坐标范围的连续轨迹点序列中的第一个和最后一个轨迹点进行标记, 通过判断对行程的起讫点进行划分。去除行程轨迹点不超过 360 个(即 30 min)的轨迹序列, 最终得到了 1 128 条有效的路径。

根据前文提出的基于偏好的路径聚类和上下文聚类方法分别获得路径的偏好类别和上下文类别。本文通过遍历簇数, 计算各簇数所对应的轮廓系数, 并选择轮廓系数最大时对应的聚类簇数。将 DDM-K-means 算法聚类簇数  $M$  设定为 3(综合最优、安全性较优、经济性较优); 将 Rock 算法聚类簇数  $N$  设定为 7, 包括单一上下文背景(4 种)和其他(3 种)。

根据上述聚类结果可以形成用户-项目交互矩阵, 构造数据集。训练集和测试集均由危货道路运输企业代号、路径类别代号和两者的交互行为构成。

### 4.2 评价指标

平均绝对百分比误差(mean absolute percentage error, MAPE)可以用来衡量预测值与真实值之间的相对偏差, 常用来作为机器学习模型预测结果好坏的衡量标准。MAPE 值越小, 模型预测准确率越高。因此本文以平均绝对百分比误差作为评价指标, 计算公式为

$$\text{MAPE} = (1/N) \sum_{v=1}^N |(y_v - \hat{y}_v)/y_v| \times 100\% \quad (6)$$

式中:  $y_v$  为用户-项目交互矩阵中的第  $v$  个元素的真实值,  $\hat{y}_v$  为用户-项目预测评分矩阵中第  $v$  个元素的预测值,  $N$  表示用户-项目交互矩阵中非空缺元素的个数。

### 4.3 对比模型

为验证本文所采用路径偏好聚类算法、路径上下文聚类算法以及推荐算法的有效性, 分别将其与两个基线模型进行对比。

路径偏好聚类算法: 1) K-means++ 算法, 该算法是对传统 K-means 聚类算法的改进, 依据最远法则选取初始聚类中心, 采用欧式距离作为相似性度量以实现聚类<sup>[13]</sup>; 2) Rock 算法, 该算法为凝聚型的层次聚类算法, 定义链接数作为相似性度量, 遵循簇间的链接数量最小, 簇内的链接数量最大的原则<sup>[14]</sup>。

路径上下文聚类算法: 1) K-means++ 算法, 该算法是对传统 K-means 聚类算法的改进, 依据最远法则选取初始聚类中心, 采用欧式距离作为相似性度量以实现聚类<sup>[13]</sup>; 2) DDM-K-means 聚类算法, 由本文提出, 该算法综合考虑了向量间的距离和方向相似性, 采用同倾向性权重加权马氏距离作为相似性度量。

推荐算法: 1) MF 算法, 该算法是经典的基于矩阵分解的协同过滤算法, 将评分矩阵分解为用户矩阵和项目矩阵, 可补全评分矩阵中的缺失值, 从而根据预测值向用户推荐<sup>[15]</sup>; 2) PMF 算法, 该算法是基于 MF 算法的进一步优化, 引入概率模型, 假设用户和项目的特征矩阵均服从高斯分布<sup>[16]</sup>。

### 4.4 实验结果与分析

从历史路径数据集中分别抽取 30%、60% 和 100% 的数据构造新的数据集, 分别表示为 DATA1、DATA2、DATA3。

在进行路径偏好聚类时, 将 K-means++、Rock 算法与本文提出的 DDM-K-means 算法在 DATA3 数据集上进行对比实验, 结果表明使用 DDM-K-means 算法进行路径偏好聚类的路径推荐平均绝对百分比误差与 K-means++ 算法相比降低了 1.24%, 与

Rock 算法相比降低了 3.65%。这表明,本文所提出的 DDM-K-means 算法相比其他聚类算法能够有效地划分路径偏好类别,进而提高运输路径推荐的准确率。

在进行路径上下文聚类时,将 K-means++、DDM-K-means 算法与本文所采用的 ROCK 算法在 DATA3 数据集上进行对比实验,结果表明与 K-means++ 算法和 DDM-K-means 算法相比,使用 ROCK 算法进行路径上下文聚类的路径推荐平均绝对百分比误差均较低。这表明,ROCK 算法更适用于本文类别型的上下文数据,聚类效果更好。

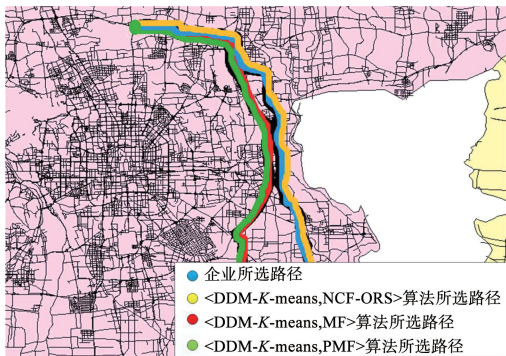
在进行路径推荐时,将 MF、PMF 算法与本文提出的 NCF-ORS 算法分别在 DATA1、DATA2、DATA3 数据集上进行对比实验,不同算法的对比结果见表 1。从表 1 可以看出,无论在多少大小比例的数据集上,使用 NCF-ORS 算法实现路径选择的准确率更高,效果更好。算法的部分路径推荐结果如图 3 所示,图中统一使用绿色圆点标记起讫点,蓝色表示危货道路运输企业所创建的真实路径,黄色、红色、

绿色分别表示 < DDM-K-means, NCF-ORS > 算法、< DDM-K-means, MF > 算法和 < DDM-K-means, PMF > 算法所推荐的路径。结果表明,NCF-ORS 算法所推荐的路径与企业真实选择一致性较高,更符合危货道路运输企业的偏好。

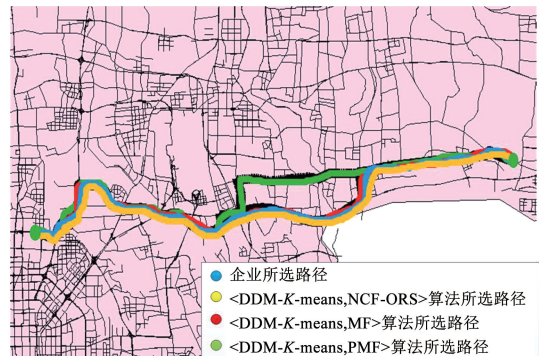
表 1 推荐算法在数据集上的实验对比结果

Tab. 1 Experimental comparative results of the recommendation algorithms on the dataset

数据集	算法	MAPE/%
DATA1	< DDM-K-means, NCF-ORS >	17.69
	< DDM-K-means, MF >	27.07
	< DDM-K-means, PMF >	25.19
DATA2	< DDM-K-means, NCF-ORS >	17.28
	< DDM-K-means, MF >	25.35
	< DDM-K-means, PMF >	24.55
DATA3	< DDM-K-means, NCF-ORS >	17.04
	< DDM-K-means, MF >	24.48
	< DDM-K-means, PMF >	23.27



(a) 路径1



(b) 路径2

图 3 路径推荐实例

Fig. 3 Illustrations of route recommendation

## 5 结 论

1) 提出了一种改进的 K-means 聚类算法 DDM-K-means,用于划分路径的偏好类别;并采用 ROCK 聚类算法划分路径的上下文类别;在此基础上,进一步提出了基于神经协同过滤的路径选择优化算法 NCF-ORS,挖掘危货道路运输企业选择路径的潜在偏好,最终构成了本文的危险货物道路运输个性化路径推荐方法。

2) 与其他聚类算法相比,本文所提出的 DDM-K-means 算法的平均绝对百分比误差最低,表明其能够有效地对路径偏好向量进行相似性度量,其聚

类结果用于路径推荐的准确率更高。

3) NCF-ORS 算法与 MF 算法、PMF 算法相比,路径推荐的平均绝对百分比误差值有所下降,不同数据集下均表现出较好性能。这表明本文所提出的路径选择优化算法能够更有效地从轨迹数据中挖掘出企业-用户之间的潜在交互信息。

4) 总体而言,本文通过挖掘轨迹数据获取路径特征,结合运输任务起讫点和上下文信息,分析企业选择路径的偏好,所提出的危险货物道路运输个性化路径推荐方法 < DDM-K-means, NCF-ORS >,能够解决特定运输任务下危货道路运输企业的路径优选问题,有助于建立科学规范的危货道路运输路径

决策体系。下一步工作将结合警情信息,评估各路径类别的风险,标记其风险等级,为危货道路运输企业推荐路径时规避风险较大的运输路径。

## 参考文献

- [1] 代存杰, 李引珍, 马昌喜, 等. 考虑风险分布特征的危险品运输路径优化[J]. 中国公路学报, 2018, 31(4): 330  
DAI Cunjie, LI Yinzhen, MA Changxi, et al. Transportation path optimization for hazardous materials considering characteristics of risk distribution[J]. China Journal of Highway and Transport, 2018, 31(4): 330. DOI:10.19721/j.cnki.1001-7372.2018.04.038
- [2] BIYELA P, RAWATLAL R. Development of an optimal state transition graph for trajectory optimisation of dynamic systems by application of Dijkstra's algorithm[J]. Computers & Chemical Engineering, 2019, 125: 569
- [3] 郭保青, 郝树运, 朱力强, 等. 基于改进蚁群算法的多 AGV 泊车路径规划[J]. 交通运输系统工程与信息, 2018, 18(6): 55  
GUO Baoqing, HAO Shuyun, ZHU Liqiang, et al. Multi-AGV parking path planning based on improved ant colony algorithm[J]. Journal of Transportation Systems Engineering and Information Technology, 2018, 18(6): 55. DOI:10.16097/j.cnki.1009-6744.2018.06.009
- [4] LIU Shan, JIANG Hai. Personalized route recommendation for ride-hailing with deep inverse reinforcement learning and real-time traffic conditions[J]. Transportation Research Part E: Logistics and Transportation Review, 2022, 164: 1. DOI:10.1016/j.tre.2022.102780
- [5] RUDLOFF C, LEODOLTER M. Personalization of routing services: user classification as a method to overcome the cold start problem for personalized models[J]. Transportation Research Record: Journal of the Transportation Research Board, 2017, 2666:38
- [6] UNNIKRISHNAN G, MATHEW D, JOSE B A, et al. Hybrid route recommender system for smarter logistics[C]// The 5th IEEE International Conference on Big Data Security on Cloud (BigDataSecurity 2019). Washington DC: IEEE Computer Society, 2019: 239
- [7] GUO Chenjuan, YANG Bin, HU Jilin, et al. Context-aware, preference-based vehicle routing[J]. VLDB Journal, 2020, 29(5): 1149
- [8] WANG Hongtao, WANG Hongmei, YI Feng, et al. Context-aware personalized path inference from large-scale GPS snippets[J]. Expert Systems with Applications, 2018, 91: 78
- [9] HE Xiangnan, LIAO Lizi, ZHANG Hanwang, et al. Neural collaborative filtering[C]// Proceedings of the 26th International Conference on World Wide Web (WWW'17). New York: ASSOC Computing Machinery, 2017: 173
- [10] 曹端喜. 聚类算法的改进和聚类有效性指标的研究[D]. 南京: 南京邮电大学, 2021  
CAO Duanxi. Improving of clustering algorithm and research on clustering validity index[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2021
- [11] ZHAO Xuemei, LI Yu, ZHAO Quanhua. Mahalanobis distance based on fuzzy clustering algorithm for image segmentation[J]. Digital Signal Processing, 2015, 43: 8
- [12] 王彬宇, 刘文芬, 胡学先, 等. 基于余弦距离选取初始簇中心的文本聚类研究[J]. 计算机工程与应用, 2018, 54(10): 11  
WANG Binyu, LIU Wenfen, HU Xuexian. Research on text clustering for selecting initial cluster center based on cosine distance[J]. Computer Engineering and Applications, 2018, 54(10): 11
- [13] 王博文, 王景升, 朱茵, 等. 基于 K-means++ 的驾驶人致因因素倾向性分析[J]. 科学技术与工程, 2021, 21(36): 15695  
WANG Bowen, WANG Jingsheng, ZHU Yin. Analysis of driver causative factor tendency based on K-means++[J]. Science Technology and Engineering, 2021, 21(36): 15695
- [14] NAOUALI S, BEN S S, CHTOUROU Z. Clustering categorical data: a survey[J]. International Journal of Information Technology & Decision Making, 2020, 19(1): 49
- [15] RENDLE S. Factorization machines with libFM[J]. ACM Transactions on Intelligent Systems and Technology, 2012, 3(3): 1
- [16] MA Hao, LIU Chao, KING I, et al. Probabilistic factor models for web site recommendation[C]// Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'11). New York: Assoc Computing Machinery, 2011: 265