

基于模糊C均值聚类的时序交通流量预测

戈 愿*, 孙德山

辽宁师范大学数学学院, 辽宁 大连

收稿日期: 2026年4月15日; 录用日期: 2026年5月8日; 发布日期: 2026年5月19日

摘 要

为了提高交通流量预测精度, 本文提出了一种模糊C均值聚类(FCM)与增强粒子群-灰狼混合优化算法(PSO-GWO)融合的交通流量预测模型。该模型通过FCM算法将交通流量划分为不同类别, 设计PSO-GWO混合优化器动态调整轻梯度提升机(LightGBM)的参数, 结合时序特征提取与选择, 构建多模式预测模型, 并应用于PEMS04单传感器交通数据集。实验结果表明, 本文所提出的FCM-PSO-GWO-LightGBM模型平均绝对误差(MAE)为9.22, 相比传统随机森林(RF)、极端梯度提升(XGBoost)、基准LightGBM模型都具有显著优势。

关键词

交通流预测, 模糊C均值聚类, 粒子群优化, 灰狼优化

Time Series Traffic Flow Prediction Based on Fuzzy C-Means Clustering

Yuan Ge*, Deshan Sun

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: April 15, 2026; accepted: May 8, 2026; published: May 19, 2026

Abstract

To improve the accuracy of traffic flow prediction, a traffic flow forecasting model integrating Fuzzy C-Means (FCM) clustering with an enhanced Particle Swarm Optimization-Grey Wolf Optimizer (PSO-GWO) hybrid algorithm is proposed. In this model, the FCM algorithm is employed to categorize traffic flow data into distinct classes. A PSO-GWO hybrid optimizer is designed to dynamically adjust the parameters of the Light Gradient Boosting Machine (LightGBM). By incorporating temporal feature extraction and selection, a multi-mode prediction model is constructed and subsequently evaluated

*通讯作者。

on the PEMS04 single-sensor traffic dataset. Experimental results demonstrate that the proposed FCM-PSO-GWO-LightGBM model achieves a Mean Absolute Error (MAE) of 9.22, exhibiting significant performance advantages over traditional models including Random Forest (RF), Extreme Gradient Boosting (XGBoost), and the baseline LightGBM model.

Keywords

Traffic Flow Prediction, Fuzzy C-Means Clustering, Particle Swarm Optimization, Grey Wolf Optimizer

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着科技的进步,我国机动车保有量持续增长,人们需求量也逐步提高。城市交通道路拥堵已经成为阻碍城市经济发展的重要因素。在此背景下智能交通系统(Intelligent transportation systems, ITS)的提出为交通流量预测技术提供了强有力的支撑[1]。通过数学模型与算法的建立,对特定路段实施交通流量监控,可以识别城市拥堵道路,并帮助相关部门提前布控,解决交通拥堵问题[2]。

然而,交通流量数据具有显著的随机性、周期性与趋势性。其变化过程不仅具有强烈的非线性,还容易受到外界因素干扰。传统的预测方法,难以捕捉其复杂的非线性特征。为了提高预测精度,许多研究者在交通流量预测方面都进行了深入的研究,同时证明了机器学习在交通流量预测方面的可行性。

张生瑞[3]等人提出了 FCM-RBF 融合的短时交通拥堵状态预测模型,将 RBF 神经网络预测得到的短时交通流参数代入 FCM 聚类结果中,取得了优异的结果。孙经伟[4]等人对 FCM 和 XGBoost 分别进行了贝叶斯优化和粒子群优化,能够稳定地识别交通状态。孙晓聪[5]等人采用随机森林对影响因素进行选择而后作用于 BLSTM 神经网络并与 ARIMA 模型结合,改善了交通流量预测精度。徐磊[6]等人基于 SSA-LightGBM 对韩城高速交通流量进行预测,展现出显著优势。

为了进一步提高预测精度,同时缓解模型对参数的依赖性。本文提出了一种混合机制预测方法。首先,采用 FCM 对交通流量数据,按照流量进行模糊聚类,探索流量数据中的内在结构,再利用 PSO 和 GWO 混合优化机制,对 LightGBM 模型进行参数优化,并与其他模型进行对比,验证本文模型的可行性。

2. 基础理论研究

2.1. 模糊 C 均值聚类算法(FCM)

交通流量数据具有明显的时序性,传统的硬聚类如(K-means)将每一个数据点固定的分一类,忽略了交通流量数据可能具有突发性,所以本文采取 FCM 聚类方法,其核心思想是,允许每个数据点以不同的隶属度属于多个聚类中心,克服了交通事故、天气等因素导致交通状态的不确定性[7]。设交通流量数据集为 $X = \{x_1, x_2, \dots, x_n\}$, n 为样本数目,为了满足数据点到聚类中心的距离尽可能小且隶属度合理,因此构造了如下的目标函数以及约束条件为:

$$J_m(U, V) = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m \|x_i - v_j\|^2 \quad (1)$$

$$\begin{cases} \sum_{j=1}^c u_{ij} = 1 \\ 0 < \sum_{i=1}^n u_{ij} < n \\ u_{ij} \in [0,1] \end{cases} \quad (2)$$

其中, c 为聚类数目, u_{ij} 是样本 x_i 对于第 j 类的隶属度; v_j 为第 j 类聚类中心; $\|x_i - v_j\|$ 为样本 x_i 到第 j 类聚类中心的欧式距离。

FCM 聚类的具体流程如下[8]:

步骤一: 初始化。输入数据集, 设置聚类数目 c , 模糊参数 m , 最大迭代次数 T_{\max} , 停止阈值 Θ 。

步骤二: 计算隶属度。按照公式(3)计算隶属度。

$$u_{ij}^{(t+1)} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - v_j^{(t+1)}\|}{\|x_i - v_k^{(t+1)}\|} \right)^{\frac{2}{m-1}}} \quad (3)$$

步骤三: 计算聚类中心, 按照公式(4)更新聚类中心。

$$v_j^{(t+1)} = \frac{\sum_{i=1}^n (u_{ij}^{(t)})^m x_i}{\sum_{i=1}^n (u_{ij}^{(t)})^m} \quad (4)$$

步骤四: 计算变化量。其中隶属度变化量如公式(5)所示:

$$\Delta_U^{t+1} = \sqrt{\sum_{i=1}^n \sum_{j=1}^c (u_{ij}^{(t+1)} - u_{ij}^{(t)})^2} \quad (5)$$

聚类中心变化量如公式(6)所示:

$$\Delta_V^{t+1} = \sqrt{\sum_{j=1}^c \sum_{k=1}^d (v_{jk}^{(t+1)} - v_{jk}^{(t)})^2} \quad (6)$$

目标函数变化量如公式(7)所示:

$$\Delta_J^{t+1} = |J_m^{(t+1)} - J_m^{(t)}| \quad (7)$$

步骤五: 终止条件判断。取最大变化量

$$\Delta = \max(\Delta_U, \Delta_V, \Delta_J) \quad (8)$$

如果 $\Delta < \varepsilon$ 或 $t > T_{\max}$ 时, 停止迭代, 输出聚类结果, 否则返回步骤二。

2.2. 粒子群算法

粒子群优化算法(Particle Swarm Optimization, PSO), 主要模拟鸟群、鱼群等生物群体觅食行为。粒子群通过分布式协作和信息共享的思想寻求全局最优解。在这个过程中粒子不仅具有个体记忆记录自身最优解, 也具有集体记忆记录整个群体最优解。如果当前位置比个体记忆优, 则更新个体记忆; 如果个体记忆比集体记忆优, 则更新集体记忆; 所有粒子都通过最优的集体记忆, 按照公式(9)、公式(10)更新速度和位置

$$h_{id}^{t+1} = w \cdot h_{id}^t + c_1 r_1 (p_{id} - s_{id}^t) + c_2 r_2 (p_{gd} - s_{id}^t) \quad (9)$$

$$s_{id}^{t+1} = s_{id}^t + h_{id}^{t+1} \quad (10)$$

其中, d 是维度, t 是当前迭代次数, w 是惯性权重, c_1 和 c_2 是学习因子, r_1 和 r_2 是 $[0,1]$ 之间的随机数, h_i 是粒子当前速度, s_i 是粒子当前位置, p_i 是个体历史最佳位置, p_g 是群体全局最佳位置。

2.3. 灰狼算法

灰狼优化算法由 Seyedali Mirjalili 于 2014 年提出, 其灵感来源于灰狼群体的严格社会等级制度和协作狩猎策略。由 α 狼(Alpha)作为领导者提供最优解; β 狼(Beta)辅助 α 狼, 监督下层, 提供次优解; δ 狼(Delta)服从 α 狼和 β 狼的命令, 并支配 ω 狼, 提供第三优解; ω 狼(Omega)服从上层所有狼, 提供其他解。灰狼算法不是单一的个体指导, 而是由三个优秀个体共同指导整个群体。其位置更新公式、距离计算公式和最终位置如(11)~(13)所示:

$$X(t+1) = X_p(t) - A \cdot D \quad (11)$$

$$D = |C \cdot X_p(t) - X(t)| \quad (12)$$

$$X_p(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (13)$$

其中 X 为灰狼当前位置, X_p 猎物位置(即 α 、 β 、 δ 的位置), t 为当前迭代次数, A 、 C 表示系数向量。系数向量是算法的关键控制参数, 具体公式如(14)~(16)所示:

$$A = 2a \cdot r_1 - a \quad (14)$$

$$C = 2 \cdot r_2 \quad (15)$$

$$a = 2 - \frac{2t}{T_{\max}} \quad (16)$$

其中 r_1 , r_2 为 $[0,1]$ 间的随机向量, T_{\max} 为最大迭代次数, a 按照公式从 2 线性递减到 0, 当 $|A| > 1$ 属于全局搜索阶段, 灰狼分散搜索, 避免陷入局部最优, 当 $|A| < 1$ 属于局部搜索阶段, 灰狼聚集在最优解周围, 进行精细搜索。

2.4. 混合机制 PSO-GWO

PSO-GWO 融合算法具体步骤如下:

步骤一: 初始化种群速度, 最优位置。

步骤二: 计算粒子适应度。

步骤三: 根据适应度排序选择 α 、 β 、 δ 狼。

步骤四: 更新 GWO 收敛因子 a , 更新 PSO 惯性权重 w , 学习因子 c_1 和 c_2 。

步骤五: 混合权重确定。权重随迭代进度线性变化, 前期 GWO 主导探索, 后期 PSO 主导开发。具体权重计算公式如(17)、(18)所示

$$\omega_g = 0.7 - 0.4 \cdot \frac{t}{T} \quad (17)$$

$$\omega_p = 1 - \omega_g \quad (18)$$

其中 ω_g 为 GWO 权重, ω_p 为 PSO 权重, t 为当前迭代次数, T 为迭代总次数。

步骤六: 对于每个非领导狼粒子, 进行 PSO 和 GWO 更新, 将两种算法结果按照步骤五中的权重进行混合。

步骤七: 判断所有粒子是否更新完成, 是, 则重新计算所有粒子适应度, 更新每个粒子的个体历史最优, 如果出现更优粒子, 更新 α 、 β 、 δ 狼; 否, 则返回步骤六。

步骤八: 每迭代 5 次, 增加随机扰动, 避免陷入局部最优。检查终止条件, 满足则输出最优解, 不满足则返回步骤四。具体流程如图 1 所示。

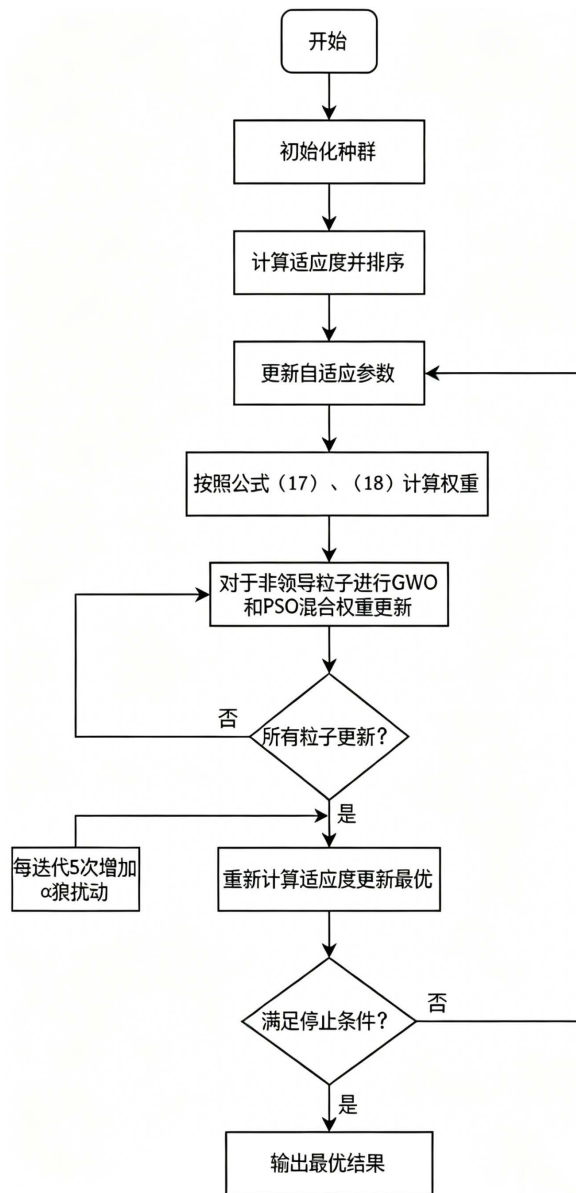


Figure 1. Flowchart of the PSO-GWO algorithm

图 1. PSO-GWO 算法流程图

3. FCM-PSO-GWO-LightGBM

LightGBM 建立在梯度提升决策树(GBDT)的基础上, 通过直方图算法降低计算的复杂度, 但是参数

众多且敏感。故本文提出了 PSO-GWO 混合优化算法对 LightGBM 进行调参, 同时考虑到交通流量的数据特征引入 FCM 算法。具体算法步骤如下:

步骤一: 数据预处理。加载数据, 进行数据清洗, 按照 7:1:2 分割为训练集、验证集、测试集, 建立滑动窗口, 每个序列包含 12 个历史时间步, 完成特征工程。

步骤二: FCM 聚类。基于交通流量特征进行 FCM 聚类, 输出聚类中心、隶属度矩阵。

步骤三: PSO-GWO 混合优化。对 FCM 聚类后的每一类都分别训练 PSO-GWO-LightGBM。

步骤四: 融合预测。采用滚动预测机制, 并判断是否需要加权预测。当前隶属度大于 0.8 时使用对应聚类模型, 否则按照隶属度进行多类加权预测。

步骤五: 模型对比。进行 FCM-PSO-GWO-LightGBM、PSO-GWO-LightGBM、LightGBM、XGBoost、RF 模型对比输出预测结果、评价指标。

全文算法流程如图 2 所示。

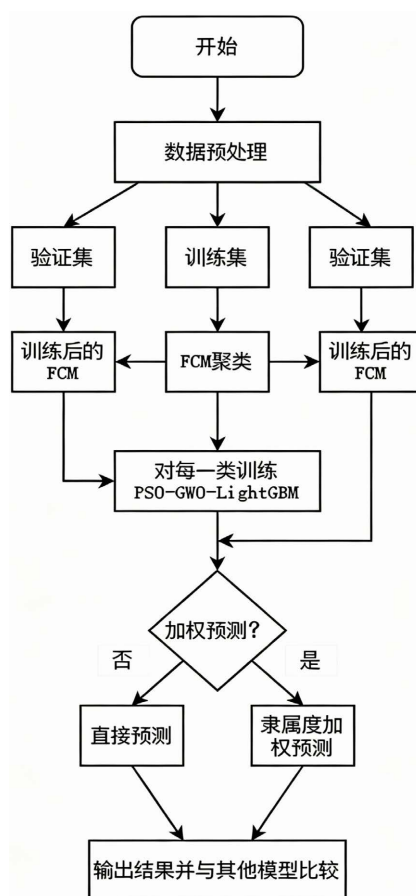


Figure 2. Flowchart of the FCM-PSO-GWO-LightGBM algorithm

图 2. FCM-PSO-GWO-LightGBM 流程图

4. 实验结果与分析

4.1. 数据来源

本文选取了 PeMS (Performance Measurement System) 系统数据作为实验数据。其中 PeMS04 数据集

是从 2018 年 1 月 1 日开始每隔 5 分钟共采集 59 天的加州第 4 区的真实交通流量数据。包括 307 个交通检测器, 涵盖流量、速度、占有率三项交通指标, 能够全面的反应交通流量数据的规律。本文研究模型聚焦时序预测能力, 剥离多传感器间的空间依赖干扰, 故选择索引为 0 的一个交通检测器用于后续实验。

4.2. 评价指标

本文采用 4 个评价指标来衡量预测性能, 分别为平均绝对误差 MAE、平均绝对百分比误差 MAPE、均方根误差 RMSE、相关系数 R^2 , 具体公式如下所示:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (19)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (20)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (21)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (22)$$

其中, n 为样本数量, y_i 为实际值, \hat{y}_i 为预测值, \bar{y} 为均值, MAE、RMSE、MAPE 的值越小, 证明结果效果越好, R^2 的值越接近于 1, 则证明拟合效果越好。

4.3. 实验结果

为了更好的验证 FCM-PSO-GWO-LightGBM 的性能, 本文引入了 XGBoost 模型、RF 模型、LightGBM 模型和 PSO-GWO-LightGBM 模型进行对比。其中 XGBoost 和 RF 模型是传统的机器学习基线模型、其余模型对比验证 PSO-GWO 优化器和 FCM 的可行性。各模型评价指标如表 1 所示。

Table 1. Evaluation metrics of each model

表 1. 各模型评价指标

| 模型 | MAE | RMSE | MAPE | R^2 |
|----------------------|-------|-------|-------|--------|
| XGBoost | 10.74 | 15.11 | 4.84% | 0.9935 |
| RF | 10.65 | 15.03 | 4.75% | 0.9936 |
| LightGBM | 10.57 | 14.81 | 4.78% | 0.9937 |
| PSO-GWO-LightGBM | 10.50 | 14.77 | 4.72% | 0.9938 |
| FCM-PSO-GWO-LightGBM | 9.22 | 13.04 | 4.17% | 0.9951 |

根据表中数据可知, 引入 PSO-GWO 混合算法对 LightGBM 的超参数进行调优后, 模型在所有指标上均有所改善, 说明智能优化算法能够有效搜索超参数空间。进一步加入 FCM 聚类后, 模型性能获得显著提升: MAE 从 10.50 降至 9.22, RMSE 从 14.77 降至 13.04, MAPE 从 4.72% 降至 4.17%。这一提升并非简单的“数据划分”所能解释, 其深层原因在于: FCM 通过模糊聚类将原始数据划分为具有不同模式特征子空间(如高峰期、平峰期等), 使 PSO-GWO 算法能够在每一类子空间内针对性地搜索与该模式相匹配的超参数。这种机制有效避免了不同模式之间的相互干扰——例如, 高峰期样本不再被平峰期数据“平均化”拖累预测值, 从而显著降低了极端误差和整体偏差。

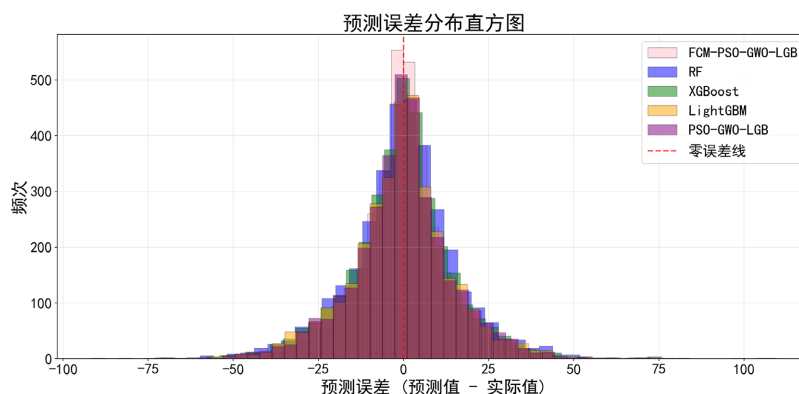


Figure 3. Histogram of prediction error distribution

图 3. 预测误差分布直方图

根据图 3 所示, FCM-PSO-GWO-LGB 模型的误差集中分布在零线附近, 峰值最高且分布区间最窄, 同时正负误差基本对称、尾部异常值最少。相比之下, RF、XGBoost、LightGBM 及 PSO-GWO-LGB 模型的误差分布更为分散。充分说明该模型具有更高的预测精度、稳定性和鲁棒性。

各模型预测结果对比如图 4 所示, 本图显示了预测的前 300 步, 由图 4 可知 FCM-PSO-GWO-LightGBM 模型的预测曲线与真实值曲线拟合最好。同时为进一步揭示各模型在复杂交通模式下的预测行为差异, 选取两个具有高波动性的典型片段进行局部放大分析, 结果如图 5、图 6 所示。

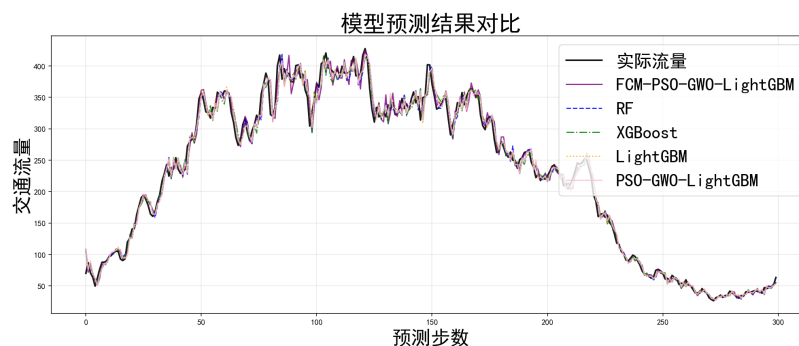


Figure 4. Comparison of model prediction results

图 4. 模型预测结果对比

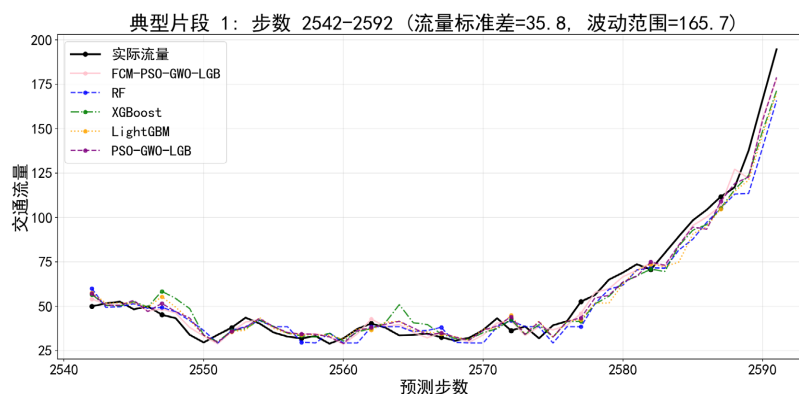


Figure 5. Model prediction fitting effect in the region of Fragment 1

图 5. 片段 1 区域的模型预测拟合效果

片段 1 的实际流量均值为 56.48, 标准差为 35.79, 表现出明显的剧变特征。在此片段中, FCM-PSO-GWO-LGB 的 MAE 为 4.10, 显著优于 RF (6.08)、XGBoost (5.79)、LightGBM (4.74) 及 PSO-GWO-LGB (4.74)。类似地, 片段 2 中, FCM-PSO-GWO-LGB 取得了最低的 MAE (2.95), 相比 PSO-GWO-LGB (3.35) 降低约 12%。上述结果表明, FCM-PSO-GWO-LGB 在高波动片段中始终表现最佳。

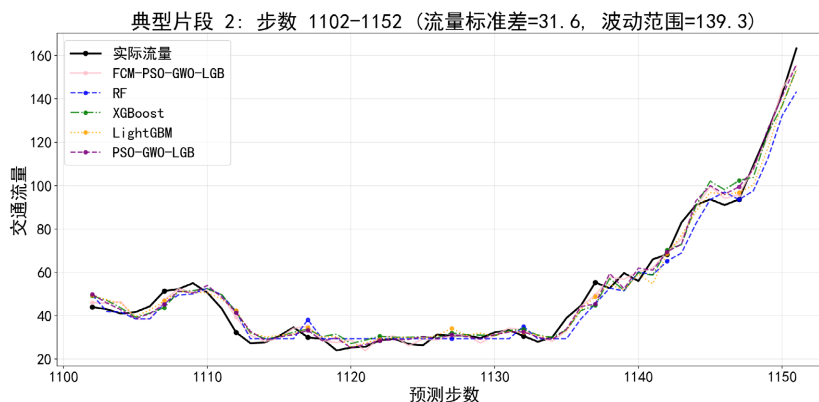


Figure 6. Model prediction fitting effect in the region of Fragment 2
图 6. 片段 2 区域的模型预测拟合效果

5. 结语

本文从经典的机器学习模型出发, 逐步引入聚类思想和智能优化算法, 提出了一种 FCM-PSO-GWO-LightGBM 模型, 通过 FCM 对数据进行软划分, 在不同的子空间实行混合优化, 最后判断是否按照隶属度进行加权预测。该模型在最后的预测阶段具有可行性, 相对于其他模型预测精度有所提升, 但是该模型还存在一些不足, 交通流量容易受到一些突发情况的影响, 比如天气、交通事故、节假日等因素, 会限制预测精度。本文模型具有多层结构, 在训练时间和资源消耗方面高于单一模型。在后续改进实验模型时, 可以考虑加入这些因素, 提高预测精度, 探索更轻量级的算法, 在精度与效率之间寻求平衡。

参考文献

- [1] 叶明壕. 基于改进 LSTM 与分层动态图卷积的交通流预测[J/OL]. 智能计算机与应用, 1-9. <https://doi.org/10.20169/j.issn.2095-2163.25112603>, 2025-12-19.
- [2] 田润智, 孙秋霞, 李琼, 等. 用于交通流量预测的动态神经张量网络模型[J]. 公路, 2025, 70(10): 257-266.
- [3] 张生瑞, 连江南, 焦帅阳, 等. 融合 FCM-RBF 的短时交通拥堵状态预测模型[J]. 重庆理工大学学报(自然科学), 2023, 37(3): 12-21.
- [4] 孙经伟, 谷远利. 基于 BO-FCM 和 PSO-XGBoost 的城市快速路交通状态识别[J]. 交通运输研究, 2023, 9(4): 64-71.
- [5] 孙晓聪, 付玉慧, 李程浩. 基于 RF-ARIMA-BLSTM 的港口船舶交通流量预测[J]. 中国航海, 2025, 48(S1): 319-326.
- [6] 徐磊, 孙朝云, 李伟, 等. 基于 SSA-LightGBM 的交通流量调查数据趋势预测[J]. 计算机系统应用, 2021, 30(1): 243-249.
- [7] 殷荣网. 基于 FCM-LSSVM 网络流量预测模型[J]. 计算机工程与应用, 2016, 52(1): 105-109.
- [8] 曹成涛, 林晓辉, 许伦辉. 联合 FCM 与群集蜘蛛优化 SVR 的短时交通流量预测[J]. 中国电子科学研究院学报, 2017, 12(1): 52-59.