

# 基于EMD-KNN-BP组合模型的高速公路短时流量预测

冯 镛<sup>1\*</sup>, 王秀兰<sup>2</sup>, 徐晓亮<sup>3</sup>, 刘兴国<sup>1</sup>, 夏传飞<sup>4</sup>

<sup>1</sup>山东交通学院交通与物流工程学院, 山东 济南

<sup>2</sup>山东公路技师学院综合办公室, 山东 济南

<sup>3</sup>山东高速集团有限公司运营管理部, 山东 济南

<sup>4</sup>临沂市交通运输执法支队, 山东 临沂

收稿日期: 2023年4月11日; 录用日期: 2023年5月24日; 发布日期: 2023年5月31日

## 摘 要

为了高效、准确地预测高速公路短时交通流量, 本文建立EMD-KNN-BP的组合模型对高速公路短时流量进行预测, 并以山东省高速公路某收费站数据进行实证分析。研究发现: 1) 基于EMD分解的组合模型能够提高短时交通流预测的精度, EMD-BP、EMD-KNN相较于单一模型的MSE、RMSE分别降低了24.4%、13.4%和9.6%、4.9%; 2) 根据IMF分量特征的不同建立的EMD-KNN-BP模型有效地降低了预测误差, 使预测效果达到最优, MSE、RMSE的误差分别降低了61.8%、38.2%。

## 关键词

短时流量预测, 经验模态分解, EMD-KNN-BP组合模型, 高速公路

# Expressway Short-Term Flow Prediction Based on EMD-KNN-BP Combined Model

Yong Feng<sup>1\*</sup>, Xiulan Wang<sup>2</sup>, Xiaoliang Xu<sup>3</sup>, Xingguo Liu<sup>1</sup>, Chuanfei Xia<sup>4</sup>

<sup>1</sup>School of Traffic and Logistics Engineering, Shandong Jiaotong University, Jinan Shandong

<sup>2</sup>General Office, Shandong Institute of Highway Technician, Jinan Shandong

<sup>3</sup>Operation Management Department, Shandong Hi-Speed Group, Jinan Shandong

<sup>4</sup>Traffic Law Enforcement Detachment, Linyi City Transport Bureau, Linyi Shandong

Received: Apr. 11<sup>th</sup>, 2023; accepted: May 24<sup>th</sup>, 2023; published: May 31<sup>st</sup>, 2023

\*通讯作者。

文章引用: 冯镛, 王秀兰, 徐晓亮, 刘兴国, 夏传飞. 基于EMD-KNN-BP组合模型的高速公路短时流量预测[J]. 交通技术, 2023, 12(3): 210-219. DOI: 10.12677/ojtt.2023.123024

## Abstract

In order to predict expressway short-term traffic flow efficiently and accurately, this paper establishes the combination model of EMD-KNN-BP to forecast expressway short-term traffic flow, and makes an empirical analysis with the data of a expressway toll station in Shandong Province. The results show that: 1) The combined model based on EMD decomposition can improve the accuracy of short-term traffic flow prediction. Compared with MSE and RMSE of single model, EMD-BP and EMD-KNN decrease by 24.4%, 13.4%, 9.6% and 4.9%, respectively. 2) The EMD-KNN-BP model established according to the different characteristics of IMF components effectively reduced the prediction error and achieved the optimal prediction effect. The errors of MSE and RMSE were reduced by 61.8% and 38.2%, respectively.

## Keywords

Short-Term Flow Prediction, Empirical Mode Decomposition, EMD-KNN-BP Combination Model, Expressway

Copyright © 2023 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

近年来,随着高速公路的迅速发展、汽车保有量快速增长,引起道路交通拥堵等问题,交通对于城市空间布局的拓展、沿线区域经济社会的发展都具有重要意义,进行高速公路短时流量预测能够及时反映高速公路的运行压力,智能交通系统就可以通过诱导系统实时发布诱导信息,提高道路交通流的通行效率,驾驶员可以根据预测结果制定出行计划,避开道路拥堵时间和出行区域,缓解交通拥堵情况。因此,如何准确高效的进行高速公路流量预测是亟待解决的问题。

目前交通流预测主要有基于统计理论、智能理论模型、非线性理论、动态交通分配、组合模型等方法。Smith等[1]、Hu等[2]分别使用BP神经网络对高速公路路段和交叉口短时交通流量进行预测,曹洁等[3]、胡伍生等[4]通过对BP神经网络进行优化以提高路段、交叉口短时流量的预测精度。祁朵和毛政元[5]、梁艳平等[6]应用了KNN算法对交叉口的数据进行预测,取得了较好的预测效果。相对与单一模型,组合模型能发挥各个模型的优点使预测结果达到最优,钱伟等[7]建立了基于灰色算法和ELM神经网络的组合模型对交叉口流量进行预测,组合模型相较于单一的ELM模型误差降低了12.8%。刘欣彤等[8]、刘钊等[9]建立了基于支持向量机和K近邻算法的组合模型对短时公交客流、交叉口交通流进行预测,仿真结果表明SVM-KNN算法可以提高预测精度。沈国江等[10]建立了基于卡尔曼滤波模型和径向基函数神经网络的组合模型对交通流进行预测,建立的组合模型相对于单一模型预测误差均有所下降。高速公路短时车流量具有非线性、随机性的特征,所以被称为非线性时间序列问题。考虑原始数据具有非平稳性,进行信号处理可以有效的提高预测精度,Ma等[11]、Chai等[12]、贺国光等[13]、曹洁等[14]建立了基于小波分解的组合模型应用于短时交通流预测,经过实证分析验证了模型的有效性,发现建立的对原始信号处理后再进行预测的组合模型可以减小误差,提高预测的精度。与小波分解相比,经验模态分

解(EMD)具有更好的适应性,吴俊和张榆锋[15]比较了经验模态分解和小波分解的特性,得出经验模态分解在处理非平稳信号的过程中具有较好的应用效果。朱从坤和谢鑫鑫[16]结合经验模态分解和 K 近邻非参数回归构建了组合模型对轨道交通站点进行预测,组合算法预测精度高于单一算法。经验模态分解相对于常见的傅里叶变换、小波变换等具有自适应性的特点,克服了傅里叶变换和小波变换的局限性,故本文选择 EMD 算法结合适用于回归分析的 K 近邻算法和自适应能力强的 BP 神经网络进行组合预测,并利用高速公路入口收费数据来检验模型的预测效果。

## 2. 基于 EMD-KNN-BP 的组合预测模型

根据不同算法的特点结合短时交通流量的特征,建立基于 EMD 分解的组合模型,可以有效地避免单一模型的缺陷,建立的组合模型具有更高的预测精度和更强的适应性。

### 2.1. 经验模态分解

经验模态分解(Empirical Mode Decomposition, EMD)是一种用于非平稳信号分析处理的方法。EMD 分解就是将复杂数据的信号进行分解,得到有限个本征模函数(Intrinsic Mode Function, IMF),分解出来的每个 IMF 分量都包含着原信号不同尺度的信息特征。

EMD 方法中分解成的 IMF 要基于以下两个条件:处理的信号在时域范围内,其极值点的个数与零点个数差小于等于 1;处理的信号上下包络线需对称于时间轴。

EMD 分解为 IMF 分量的具体步骤如下:

- 1) 找极值点。寻找原始信号  $X(t)$  的极值点,包括极大值和极小值;
- 2) 得到极大、极小值包络线。通过样条曲线连接极大值点和极小值点,得到包络  $e_{\max}(t)$  和  $e_{\min}(t)$
- 3) 计算平均包络线。公式如下:

$$m_{k,i}(t) = \frac{1}{2}(e_{\max}(t) + e_{\min}(t)) \quad (1)$$

4) 迭代。将原始信号  $X(t)$  减去平均包络线得到  $c_1$ ,若  $c_1$  不满足条件则重复上述条件直至满足条件,得到的  $c_m$  为 IMF1。将信号分离出来后多次迭代,直至剩余信号为单调函数。

### 2.2. K 近邻算法

KNN (K-Nearest Neighbor)算法,又称 K 最邻近法,是由 Cover 和 Hart 于 1968 年提出的概念。KNN 算法是一种较为成熟的机器学习算法,属于数据挖掘的分类技术,通过测量不同特征值之间的距离方法进行分类。

KNN 的算法原理是存在一个样本集合,并且样本集中的数据已被分为不同的类别,输入没有类别的新数据后,将新数据的特征与样本集中的数据特征进行比较,然后提起样本集特征的最相似的 K 个分类标签,选择 K 个标签中次数最多的分类为新数据的类别。在 KNN 模型的训练过程中, K 值是一个重要的训练参数, K 的取值受问题自身和数据集的大小决定,如果 K 值过小,容易受个例影响; K 值过大,容易受到距离较远的特殊数据影响。

本文采用欧氏距离作为样本距离,  $X_n$  为实数向量空间  $R_n$ ,  $x_i, x_j \in X_n$ ,  $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)})^T$ ,  $x_j = (x_j^{(1)}, x_j^{(2)}, \dots, x_j^{(n)})^T$ , 样本点和训练集中样本点的距离 L 的计算公式如下所示:

$$L(x_i, x_j) = \left( \sum_{l=1}^n |x_i^{(l)} - x_j^{(l)}|^2 \right)^{1/2} \quad (2)$$

### 2.3. BP 神经网络

BP (back propagation)神经网络是由科学家 Rumelhart 和 McClelland 等人于 1986 年提出的概念,是目前应用最广泛的神经网络模型之一。神经元模型是通过模仿生物神经元结构而被设计出来的,经过输入-传输-输出进行变换,基本思想是梯度下降法。BP 神经网络作为机器学习算法的一种,就是通过用机器学习算法建立模型,让机器模拟人的决策行为进行对数据进行处理,通过不断对模型进行训练,再进行了错误的预测之后对模型进行修正。

BP 神经网络主要是由输入层、隐藏层、输出层组成。输入层(Input Layer)是指外界已有信息通过输入层输入到模型中,从而传递到下一神经元;隐藏层为神经元内部的交换环节,对传入的信息进行转换,隐藏层(Hidden Layer)可以为单层或多层,每一层神经元只会对下一层的神经元产生影响;输出层(Output Layer)将隐藏层变换得到的信息进行输出,以需要的形式展现结果。BP 神经网络是一种多层前馈神经网络,它的主要特点是将信号向前进行传递,误差反向进行传递,正向传播是将输入的样本信息从输入层传入,经过隐藏层内部处理后通过输出层传出,如果输出层的输出内容与期望的输出结果不符则进入误差的反向传播。反向传播是将输出内容以某种形式通过隐藏层反向向输入层传播,将误差分配给各层中的所有单元,从而获得修正各单元的误差信号,从而使最终的输出结果不断向期望值逼近,产生的误差最小。BP 神经网络架构图如下图 1 所示:

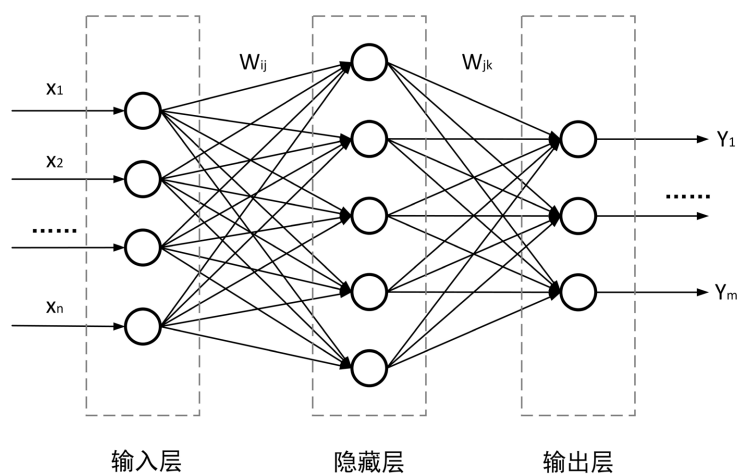


Figure 1. Architecture diagram of BP neural network

图 1. BP 神经网络架构图

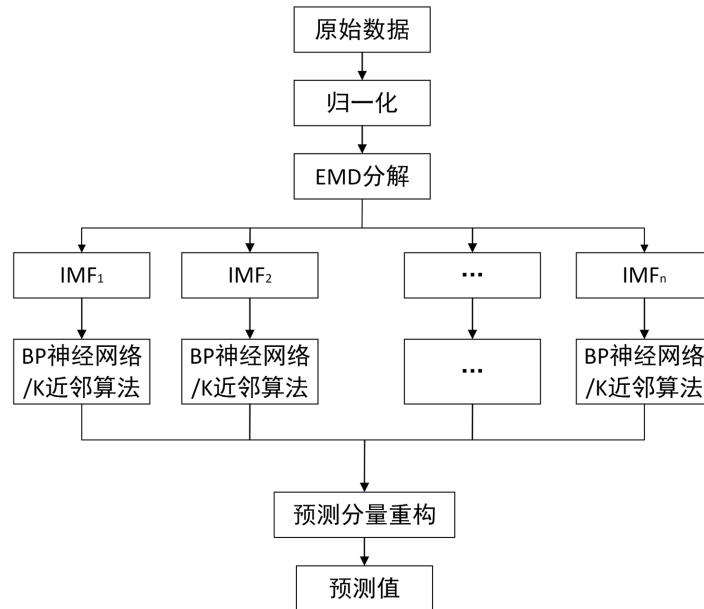
图 1 中  $x_1, x_2, \dots, x_n$  表示 BP 神经网络中的输入数据, 可以看作函数的自变量;  $W_{ij}, W_{jk}$  表示 BP 神经网络中从神经元  $i-j, j-k$  的连接权值;  $y_1, y_2, \dots, y_m$  表示 BP 神经网络中的输出数据, 可以看作函数的因变量。任何在闭区的一个连续函数都可以用一个隐层的 BP 网络逼近, 一个三层的 BP 网络可以完成任意的 N 维到 M 维的映射, 因此在 BP 网络多采用单隐层网络[17]。

### 2.4. 组合模型的构建

组合模型首先是将原始数据归一化后进行 EMD 分解, 将分解得到的 IMF 分量使用 BP 神经网络或 K 近邻算法进行预测, 预测分量重构后得到预测值。建立的组合模型如下图 2 所示。

步骤一: 对原始数据进行预处理, 将收费数据转化为随时间变化的车流量数据, 对缺失数据和错误数据进行修正, 并对数据进行标准化处理, 将处理后的数据按照 10 min 的精度进行划分, 得到数据集  $X(t)$ ,

将  $X(t)$  分为两列, 第一列为数据的前 80% 作为训练集  $X$  输入对模型进行训练, 第二列为数据剩余的 20% 为预测值。



**Figure 2.** Combined model structure diagram  
**图 2.** 组合模型结构图

步骤二: 对数据进行归一化处理, 公式如下所示:

$$Z = (X - X_{\min}) / (X_{\max} - X_{\min}) \quad (3)$$

式(3)中:  $Z$  为归一化后的数值;  $X$  为样本数据;  $X_{\max}$  为样本数据的最大值;  $X_{\min}$  为样本数据的最小值。

步骤三: 将时间序列  $X(t)$  进行 EMD 分解, 将得到  $n$  个 IMF 序列  $c_i(t) (i=1, 2, \dots, n)$ , 即:

$$X(t) = \sum_{i=1}^n c_i \quad (4)$$

步骤四: 首先需要确定  $K$  的值, 结合理论并通过枚举法发现, 当  $K$  为 8 时算法的各项评价指标达到最优, 将 EMD 分解得到的高频 IMF 序列输入 KNN 算法中。

步骤五: 基于 BP 神经网络的相关理论, 选用单隐层网络, 学习率 0.02, 迭代次数 1000 次, 将低频序列输入 BP 神经网络中。

步骤六: 对输出结果进行反归一化处理, 公式如下所示:

$$X = Z(X_{\max} - X_{\min}) + X_{\min} \quad (5)$$

步骤七: 输出预测值  $Y_{i1}$ 、 $Y_{i2}$ , 得到最终的预测结果, 即:

$$Y = (Y_{i1} + Y_{i2}) \quad (6)$$

## 2.5. 误差评价指标

为比较不同算法的计算结果, 需要建立统一的评价指标体系进行分析, 本文选用均方根误差(RMSE)和平均绝对百分比误差(MAPE)两种常见的评价指标对预测的模型进行标准化评价。

具体计算公式如下所示:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^n (Y_i - Y'_i)^2} \quad (7)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^n \left| \frac{Y_i - Y'_i}{Y'_i} \right| \quad (8)$$

其中,  $N$  为样本数量,  $Y_i$  为实际值,  $Y'_i$  为预测值。

### 3. 实例分析

#### 3.1. 数据处理分析

实验数据来源于高速公路收费站入口处车辆的信息, 选用济南某高速公路收费站入口车辆收费数据作为测试样本进行预测, 样本记录 7 月 3 日 0 时至 7 月 7 日 24 时进入高速公路的收费数据, 按照十分钟的精度分为 720 组。高速公路入口流量具有周期性, 单日流量呈现双峰分布, 早高峰为上午 8 时, 晚高峰为下午 17 时, 下图 3 为 7 月 3 日至 7 月 7 日高速公路入口车流量变化趋势。

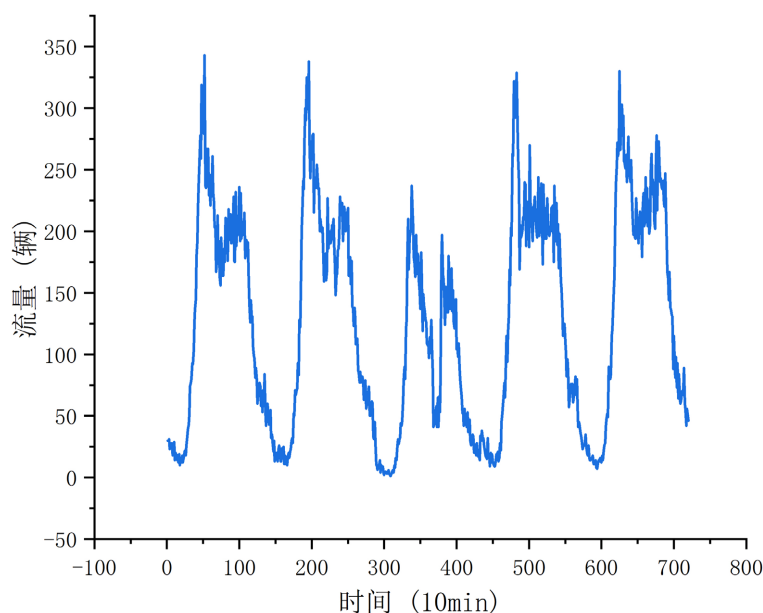


Figure 3. Trend chart of expressway inlet flow change  
图 3. 高速公路入口流量变化趋势图

#### 3.2. 基于 EMD 分解的组合模型预测

基于 python 语言, 依据 EMD 去噪算法, 对高速公路入口流量处理后的数据进行 EMD 分解。分解后得到 8 条 IMF 函数序列(IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, IMF8), 如图 4 所示。

从图中可以看出数据具有非平稳性, EMD 分解可以降低原始序列的复杂度, IMF 分量逐渐趋于平稳, 将分解得到的各个 IMF 分量分别输入到构建的 BP 神经网络和 KNN 模型中, 其中前 576 组数据作为训练数据对 EMD-KNN、EMD-BP 网络进行训练, 后 144 组数据对模型的预测精度进行检验, 根据将不同分量得到的预测结果进行重构后得到预测数据, 单一模型和组合模型的对比图如下图 5, 图 6 所示。

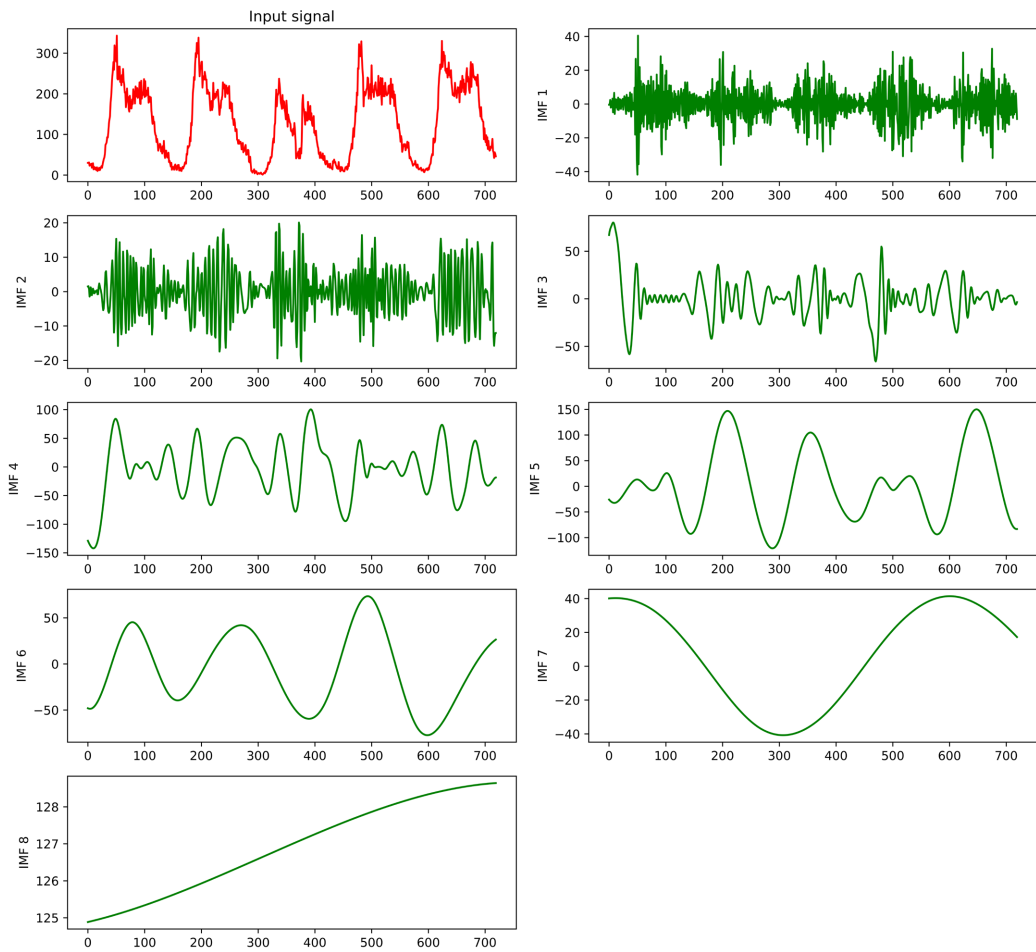


Figure 4. EMD breakdown of expressway flow  
图 4. 高速公路流量 EMD 分解图

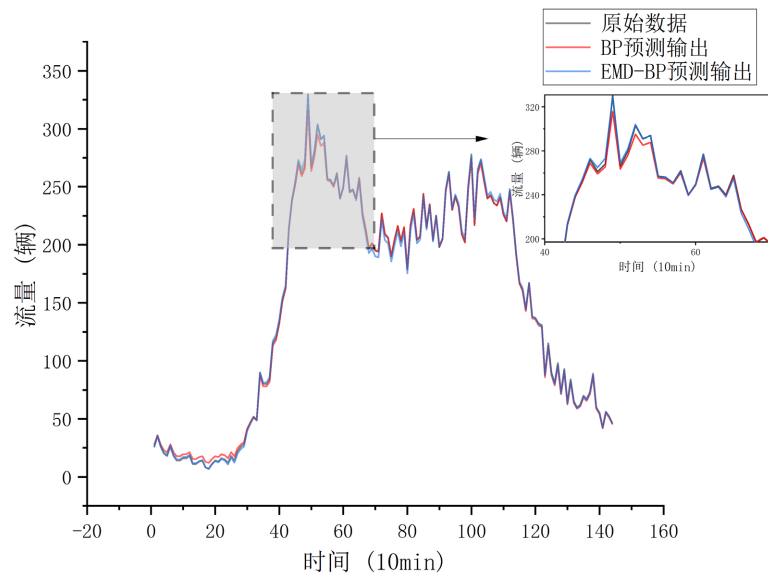


Figure 5. BP neural network, EMD-BP algorithm prediction graph  
图 5. BP 神经网络、EMD-BP 算法预测图

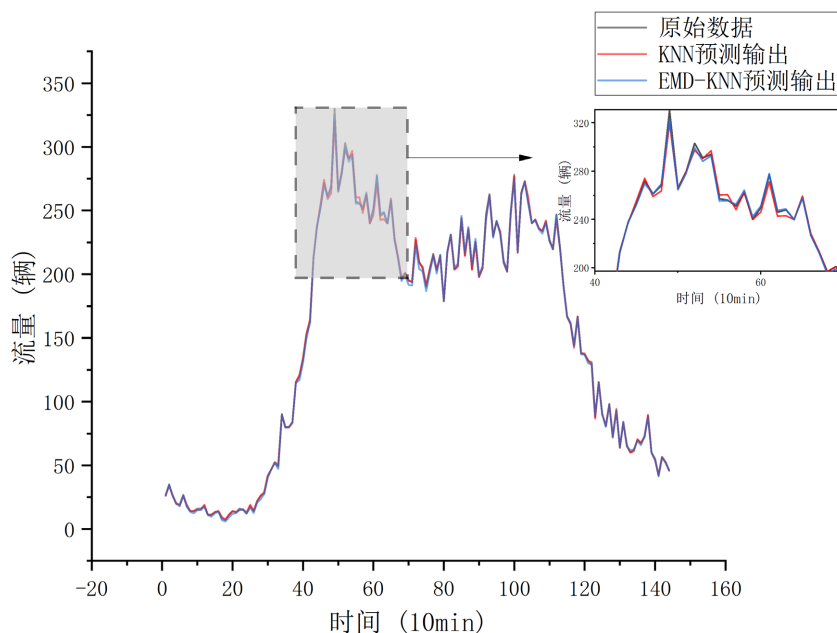


Figure 6. KNN algorithm, EMD-KNN prediction graph

图 6. KNN 算法、EMD-KNN 预测图

从对不同频率 IMF 分量的预测中可以看出, KNN 算法对于高频率序列的预测拟合效果优于 BP 神经网络, 低频序列则反之, 将高频序列(IMF1、IMF2、IMF3、IMF4)输入 KNN 模型中进行预测, 低频序列(IMF5、IMF6、IMF7、IMF8)输入 BP 神经网络进行预测, 建立的 EMD-KNN-BP 组合模型预测结果如下图 7 所示:

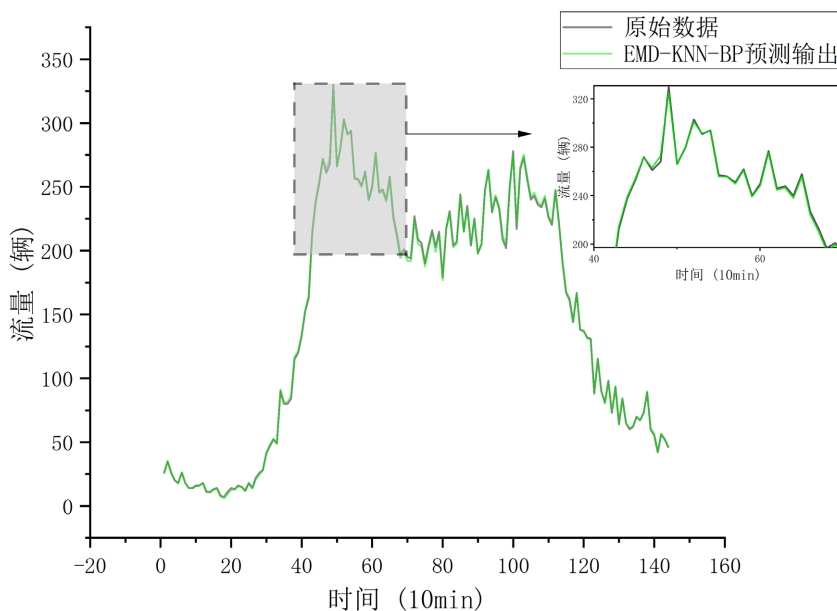


Figure 7. EMD-KNN-BP algorithm prediction graph

图 7. EMD-KNN-BP 算法预测图

从图中可以看出, 基于 BP 神经网络和 KNN 算法的预测结果和原始数据的拟合曲线相似, 但拟合效果相对较差, KNN 算法的预测结果优于 BP 神经网络, 但与真实值仍有一定差距。EMD 分解后结合 BP



神经网络和 KNN 算法的预测的结果均优于单一模型。

### 3.3. 模型评价

选择 MSE 和 RMSE 对预测结果进行评价, 得到的单一模型和组合模型的预测误差如下表 1 所示:

Table 1. Prediction model evaluation index

表 1. 预测模型评价指标

数量	BP	KNN	EMD-BP	EMD-KNN	EMD-KNN-BP
MSE	6.1641	3.0341	4.6578	2.7421	2.3552
RMSE	2.4828	1.7419	2.1582	1.6559	1.5346

从表中可以看出, BP 神经网络的预测误差最大, 利用 EMD 分解进行数据处理后的 EMD-BP 和 EMD-KNN 组合模型相较于单一模型的预测精度均有提高, 这说明 EMD 分解可以有效地提高数据质量, 提高单一模型的预测精度; 根据分量特征不同而重新组合的 EMD-KNN-BP 模型的预测误差 MSE、RMSE 最低, 具有良好的预测效果。

### 4. 结语

针对高速公路短时流量预测问题, 本文提出了引用 EMD 分解对数据进行处理, 构建组合模型解决单一模型的缺陷, 文中还选用了 BP 神经网络和 K 近邻算法进行对照, 并结合 EMD 分解不同分量的组合构建了 EMD-KNN-BP 组合模型进行预测。本文以高速公路收费站入口数据对建立的组合模型进行实证分析, 最终建立的组合模型 EMD-KNN-BP 的预测误差显著低于 BP 神经网络和 KNN 算法两种基本模型, 通过对预测结果进行分析可以得出: 经过 EMD 分解得到的数据具有更好的平稳性, 能够提高模型的预测精度, EMD-BP、EMD-KNN 相较于单一模型的 MSE、RMSE 分别降低了 24.4%、13.4%和 9.6%、4.9%; 组合模型 EMD-KNN-BP 可以有效地结合单一模型的优点, 预测效果优于 EMD-KNN 和 EMD-BP 组合模型, 有效地提高了高速公路短时流量的预测精度, MSE、RMSE 的误差分别降低了 61.8%、38.2%。

### 基金项目

山东省交通科技计划“基于高速公路优势资源和客户需求的综合服务体系构建(2020B50)”。

### 参考文献

- [1] Smith, B.L. and Demetsky, M.J. (1994) Short-Term Traffic Flow Prediction: Neural Network Approach. *Transportation Research Record*, No. 1453, 98-104.
- [2] Hu, W., Liu, Y., Li, L., et al. (2010) The Short-Term Traffic Flow Prediction Based on Neural Network. *IEEE International Conference on Future Computer and Communication*, Wuhan, 21-24 May 2010, V1-293-V1-296. <https://doi.org/10.1109/ICFCC.2010.5497785>
- [3] 曹洁, 沈钧珥, 张红, 陈作汉, 侯亮. 优化的 BP 神经网络短时交通流预测方法[J]. *传感器与微系统*, 2020, 39(5): 58-60+64.
- [4] 胡伍生, 吕楚男, 夏晓明. 基于神经网络的短期交通流预测模型[J]. *现代测绘*, 2020, 43(5): 10-13.
- [5] 祁朵, 毛政元. 基于自适应时序剖分与 KNN 的短时交通流量预测[J]. *地球信息科学学报*, 2022, 24(2): 339-351.
- [6] 梁艳平, 毛政元, 邹为彬, 许锐. 基于相似数据聚合与变 K 值 KNN 的短时交通流量预测[J]. *地球信息科学学报*, 2018, 20(10): 1403-1411.
- [7] 钱伟, 车凯, 李冰峰. 基于组合模型的短时交通流量预测[J]. *控制工程*, 2019, 26(1): 125-130.

- 
- [8] 刘欣彤, 黄小龙, 谢秉磊. 基于 SVM-KNN 的降雨条件下短时公交客流预测[J]. 交通信息与安全, 2018, 36(5): 117-123.
- [9] 刘钊, 杜威, 闫冬梅, 柴干, 郭建华. 基于 K 近邻算法和支持向量回归组合的短时交通流预测[J]. 公路交通科技, 2017, 34(5): 122-128+158.
- [10] 沈国江, 朱芸, 钱晓杰, 胡越. 短时交通流组合模型预测[J]. 南京理工大学学报, 2014, 38(2): 246-251.
- [11] Ma, L.-C., Xu, W.-L., *et al.* (2011) Prediction Model of Traffic Flow Along Typical Roads in City Urban District Based on Wavelet Transform. *Control and Decision*, **26**, 789-793.
- [12] Chai, Y.C., Huang, D.R. and Zhao, L. (2016) A Short-Term Traffic Flow Prediction Method Based on Wavelet Analysis and Neural Network. *IEEE Control & Decision Conference*, Yinchuan, 28-30 May 2016, 7030-7034. <https://doi.org/10.1109/CCDC.2016.7532264>
- [13] 贺国光, 马寿峰, 李宇. 基于小波分解与重构的交通流短时预测法[J]. 系统工程理论与实践, 2002, 22(9): 101-106.
- [14] 曹洁, 沈钧珥, 张红, 等. 基于小波和多维重构的 BP 神经网络交通流短时预测[J]. 计算机应用与软件, 2018, 35(12): 67-71+88.
- [15] 吴俊, 张榆锋. 经验模态分解和小波分解滤波特性的比较研究[J]. 云南大学学报(自然科学版), 2012, 34(3): 285-290+297.
- [16] 朱从坤, 谢鑫鑫. 基于经验模态分解与 K 近邻的轨道交通站点客流预测方法[J]. 武汉理工大学学报(交通科学与工程版), 2022, 46(6): 997-1002.
- [17] 戚德虎, 康继昌. BP 神经网络的设计[J]. 计算机工程与设计, 1998(2): 47-49.