

基于改进 KNN 算法的交通流异常数据修复方法

秦一菲, 马明辉, 王岩松, 郭辉, 张亮

(上海工程技术大学 汽车工程学院, 上海 201620)

摘要: 交通流数据分析是交通规划、控制、管理等工作实施的基础; 交通流数据异常会给交通状态辨识及交通管理和控制带来困扰, 不利于交通领域各方面研究及工作的开展; 因此, 对异常数据进行修复具有必要性; 为了提高交通流异常数据修复精度, 进一步改善交通数据质量, 构建了基于改进 K 近邻 (K-Nearest Neighbor, KNN) 算法的交通流异常数据修复模型; 通过对 KNN 基础模型中 k 值和状态向量进行优选、提出距离占比的近邻值权重选取方式, 实现对其模型的改进; 为了验证模型的有效性, 采用实测交通流数据进行实验分析; 实验结果表明, 改进的 KNN 数据修复模型具有更高的修复精度, 其平均相对误差为 9.88%, 能够有效改善数据质量, 为智能交通控制体系提供基础数据支持。

关键词: 交通流; 异常数据修复; KNN 算法; 近邻值

A Recovery Method for Abnormal Traffic Flow Data Based on Improved KNN Algorithm

Qin Yifei, Ma Minghui, Wang Yansong, Guo Hui, Zhang Liang

(School of Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: Traffic flow data analysis is the basis for implementation of traffic planning, control, and management. Abnormal traffic data which is not conducive to all aspects of transport research and related work brings difficulties to the identification of traffic conditions, traffic management and control. Therefore, it is necessary to repair abnormal data. To improve the recovery accuracy of traffic flow anomaly data and further improve the quality of traffic data, a model of traffic flow anomaly data recovery method based on improved K-Nearest Neighbor (KNN) algorithm was constructed. The model is improved by optimizing the k-values and state vectors in the KNN basic model, and proposing a distance weights as the selection of neighboring weights values. In order to verify the validity of the model, the measured traffic flow data was used for experimental analysis. The experimental results show that the improved KNN data recovery model has higher recovery accuracy, the mean average relative error is 9.88%. It can effectively improve the data quality, and provide basic data support for the intelligent traffic control system.

Keywords: traffic flow; abnormal data recovery; KNN algorithm; near neighbor

0 引言

交通数据是智能交通系统 (intelligent transportation system, ITS) 中主要的支撑, 可以直接反映交通的运行状态, 能够为交通规划、管理和决策提供重要的信息^[1]。交通流数据质量的高低直接影响着交通流状态分析的准确度。交通数据在实际获取中, 由于受到检测器自身异常、传输网络异常及环境因素等的影响, 采集到的交通流数据难免会出现数据缺失、数据错误等质量问题^[2], 对交通流

数据分析以及交通管控精准性造成影响。因此, 进行异常数据修复, 提高交通流检测数据的质量, 是交通状态辨识、交通管理及控制等工作正常进行的必然要求。

交通流异常数据修复方面的研究通常是基于历史数据构建数据修复模型^[3]。目前, 对交通流异常数据修复方法的研究大致可以分为三类: 一类是传统数学模型, 包括历史趋势法^[4]、移动平均法^[5]、插值法^[6-7]、灰色残差模型^[8]等; 第二类是智能修复方法, 包括非参数回归^[9]、神经网络^[10]等; 第三类是组合模型^[11-12], 是指两种或两种以上的数据处理方法进行组合的方法。智能修复方法主要是采用机器学习算法对异常数据进行识别和修复, 该方法是目前普遍采用的数据修复方式^[13]。

K 最近邻 (K-Nearest Neighbor, KNN) 算法模型属于非参数回归方法, 是一个理论上比较成熟的方法, 也是最基本的机器学习算法之一。在交通领域中, KNN 算法模型早期一般用于交通流预测。Davis 和 Nihan^[14]率先运用 k 近邻法的非参数理论方法对高速路交通流量进行预测。随着 KNN 模型在交通领域的引入, 有研究者针对如何提升 k 近邻运算速度的问题进行了一系列探索^[15-16]。随后, 于滨等^[17]分析了 k 近邻算法的时间和空间参数, 提出 4 种状态

收稿日期: 2018-05-18; 修回日期: 2018-06-15。

基金项目: 国家自然科学基金项目 (51675324); 上汽基金项目 (1523)。

作者简介: 秦一菲 (1993-), 女, 河南省周口市人, 硕士研究生, 主要从事智能交通控制方向的研究。

王岩松 (1971-), 男, 山东省文登市人, 硕士生导师, 教授, 主要从事车辆噪声、振动测试与控制技术 (NVH) 方向的研究。

郭辉 (1981-), 男, 山东省曹县人, 硕士生导师, 副教授, 主要从事车辆 NVH 技术、智能压电机构设计方向的研究。

通讯作者: 马明辉 (1989-), 女, 吉林省长春市人, 硕士生导师, 博士, 主要从事智能交通控制方向的研究。

向量组合的 k 近邻模型, 通过优化算法参数来提高数据处理精度。谢海红等^[18]分析了原有的短时交通流预测的 k 近邻算法, 用模式距离搜索方法代替原有的欧氏距离搜索方法, 引入多元统计回归模型, 建立了一种改进的 k 近邻算法, 结果表明, 改进的 k 近邻算法应用效果比原有 k 近邻算法要好。Habtichael F. G.^[19]通过采用加权欧氏距离作为相似性度量标准, 采用指数权重作为近邻值权重, 提出了非参数和数据驱动的短期交通流预测方法, 结果证明其方法能有效提高预测精度。

目前 KNN 算法模型主要应用于交通流预测方面, 在交通数据修复方面的应用和研究较少。因此, 本文在对 KNN 算法特性分析的基础上, 通过对 k 值、修复状态向量以及近邻值权重等模型主要相关要素的优选和改进, 进而构建基于改进 KNN 算法的交通流异常数据修复模型。研究结果旨在为交通流数据修复方法提供一种新的思路, 为交通管控提供准确的基础数据支持。

1 KNN 异常数据修复模型

1.1 KNN 基础算法

KNN 算法主要以数据为驱动, 通过搜索历史数据库中当前数据相似的状态向量数据来进行数据修复。模型的基本流程如下: 首先, 构建具有较大容量且有代表性的历史数据库。其次, 设定模型的相关要素, 包括近邻个数 k 、状态向量, 距离度量方式和修复算法等, 其中近邻个数 k 、状态向量和距离度量方式组成了模型的搜索机制。最后, 根据输入的异常数据状态向量和搜索机制, 从历史数据库中找到与异常数据状态向量相匹配的近邻, 利用修复算法得到修复后的交通数据值。

模型的运算速度和修复精度主要与 4 个要素密切相关: 近邻个数、状态向量、距离度量方式、算法实施。模型性能的优劣很大程度上取决于这 4 个要素的选取。近邻数 k 表示从历史数据库中选取与当前数据相似的数据组数, k 值的选取主要与历史数据库有关。状态向量作为当前数据与历史数据库实施匹配的一个匹配标准, 表征了历史数据库中的数据特征。距离度量方式用于计算异常数据状态向量与历史数据库中各个状态向量之间的距离。修复算法构造了采用 k 组近邻数据集修复异常数据值的方案。

1.2 改进 KNN 算法

异常数据修复是交通管控有效实施的基础, 采用 KNN 算法能够在众多的历史数据组群中匹配到异常数据相似组群, 进而实现对异常数据的修复。因此, 本文提出基于改进 KNN 算法的异常数据修复方法。KNN 异常数据修复模型实施过程如图 1 所示。

针对出现异常的交通流数据, 运用改进 KNN 算法对其异常数据进行修复, 其实现方法具体步骤如下:

1) 选取异常数据出现的同一监测点的正常交通流检测器数据作为数据源, 选取连续 5 个数据为一组, 建立历史数据状态向量库 X_n 。

2) 识别异常数据, 选取异常数据前后两个时间间隔的

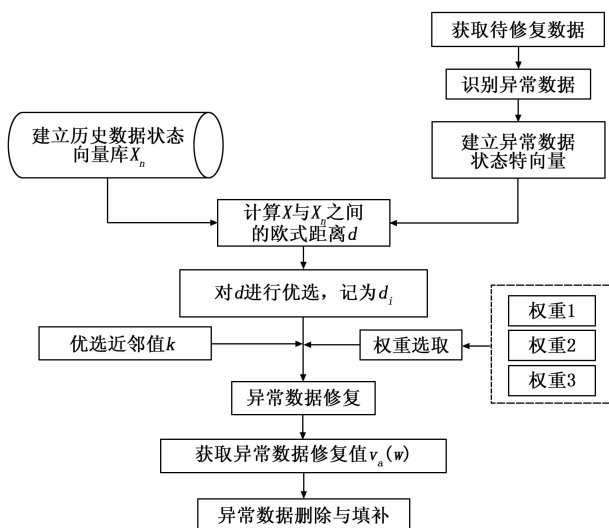


图 1 KNN 数据修复方法实施过程

正常数据, 建立异常数据状态向量 X 。

3) 计算异常数据状态向量 X 与历史数据状态向量库 X_n 之间的欧式距离 d , 对其距离按照由小到大进行排序, 记为 d_i 。

4) 通过数据验证平均相对误差与 k 值的关系, 对近邻值个数 k 进行优选。

5) 选取距离占比的近邻值权重选取方式。

6) 运用修复算法对异常数据进行修复, 得出修复后的数值 $\hat{v}_a(w)$ 。

7) 对异常值进行删除, 并填补修复后的数值。

本文在对 KNN 基础模型以及交通流数据特性分析的基础上, 对 KNN 基础模型的改进主要体现在三点: 一是对 k 值进行优选; 二是考虑异常数据前后时刻的关联数据, 建立了状态向量; 三是提出了一种新的近邻值权重选取方式。

1) 参数 k 优化选取:

近邻值个数 k 表示从历史数据库中选取的近邻个数, k 值大小是影响算法计算时间和数据修复精度的一个重要因素。一般情况下, k 值越大, 修复精度越高但其运算时间会相应增加。

由图 2 可知, 随着 k 值的增大, 平均相对误差不断减小, 并逐渐趋于平缓。当 k 值为 25 时, 图形基本收敛, 随着 k 值再增大, 平均相对误差已无明显变化。因此本文优选 k 值为 25。

2) 状态向量建立:

分别建立历史数据状态向量库和异常数据状态向量。

目前对交通流异常数据的修复, 通常考虑的是利用异常数据的前向数据对其进行修复。图 3 表示一段时间内交通流速度变化趋势。从图中可以看出, 交通流速度随着时间的变化, 具有不确定性、波动性和随机时变特性。交通流数据自身具有一定的特征和规律, 相邻时间之间的数据具有某种关联性, 前一刻速度大小会影响后一刻速度大小, 数据之间具有很强的连续性和关联性。并且时间距

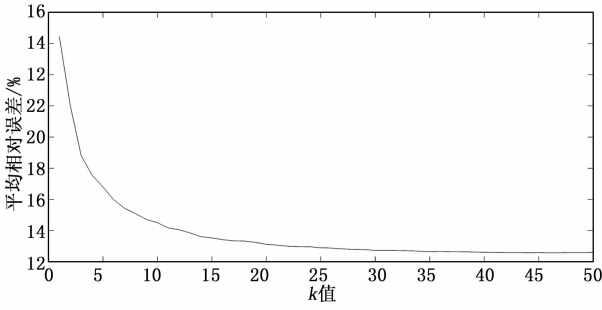


图 2 平均相对误差与 k 值的关系

离越近的数据之间关联性越大^[20]。

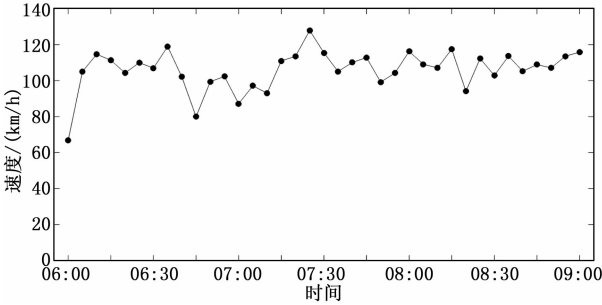


图 3 交通流速度数据趋势图

根据交通流数据之间的这种关联性，按照时间序列，建立以连续 5 个数据为一组的历史数据状态向量库 $\mathbf{X}_n, \mathbf{X}_n = \{v_{h1}^n, v_{h2}^n, v_{h3}^n, v_{h4}^n, v_{h5}^n\} (n = 1, 2, 3, \dots)$ 。按照同样的状态向量建立方式，分别选取异常数据前后两个时间间隔的正常数据，建立异常数据状态向量 $\mathbf{X}, \mathbf{X} = \{v_{a1}, v_{a2}, v_a(\omega), v_{a4}, v_{a5}\}$ ，其中 $v_a(\omega)$ 为异常数据。历史数据状态向量库可以表示为：

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} = \begin{bmatrix} v_{h1}^1 & v_{h2}^1 & v_{h3}^1 & v_{h4}^1 & v_{h5}^1 \\ v_{h1}^2 & v_{h2}^2 & v_{h3}^2 & v_{h4}^2 & v_{h5}^2 \\ v_{h1}^3 & v_{h2}^3 & v_{h3}^3 & v_{h4}^3 & v_{h5}^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_{h1}^n & v_{h2}^n & v_{h3}^n & v_{h4}^n & v_{h5}^n \end{bmatrix} \quad (1)$$

3) 距离度量：

距离度量方式是用于度量异常数据状态向量与历史数据库中各个状态向量之间的相似程度，一般采用欧氏距离作为衡量标准：

$$d_i = \sqrt{\sum_j (v_{aj} - v_{hj}^i)^2} \quad (2)$$

式中， d_i 为异常数据状态向量与历史数据库中第 i 组状态向量的距离， $i = 1, 2, 3, \dots, n$ ； v_{aj} 为异常数据状态向量中第 j 个子数据的值， $j = 1, 2, 4, 5$ ； v_{hj}^i 为第 i 组历史数据状态向量中第 j 个子数据的值。

4) 修复算法构建：

修复算法构造了采用 k 组近邻数据集修复异常数据值的方案，即：

$$\hat{v}_a(\omega) = \sum_{i=1}^k \alpha_i v_{h3}^i(\omega) \quad (3)$$

式中， $v_{h3}^i(\omega)$ 为异常数据对应的历史数据库中第 i 组状态向量数据中的子数据； α_i 为异常数据对应的历史数据库中第 i 组状

态向量数据中子数据所占的权重； $\hat{v}_a(\omega)$ 为修复后的异常数据。

寻找到的 k 个最近邻中，距离越近，说明此近邻值与当前异常数据状态向量的相似度越大。为了提高其修复精度，采取对其近邻值赋予权重的方式，距离越近的近邻值，赋予的权重越大。

关于近邻值权重的选取，现有的 k 近邻权重选取方式主要有距离倒数权重和排队指数权重两种，分别定义为权重 1、权重 2。如式 (4)、式 (5)。

距离倒数权重：

$$\alpha_i = \frac{d_i^{-1}}{\sum_{i=1}^k d_i^{-1}} \quad (4)$$

排队指数权重：

$$\alpha_i = \frac{(k - i + 1)^2}{\sum_{i=1}^k (k - i + 1)^2} \quad (5)$$

在现有权重的基础上，根据 KNN 算法的基本原理，综合考虑算法复杂度和计算速度，提出距离占比的权重选取方式，定义为权重 3，如式 (6)。

距离占比权重：

$$\alpha_i = \frac{\sum_{i=1}^k d_i - d_i}{(k - 1) \sum_{i=1}^k d_i} \quad (6)$$

式中， α_i 是第 i 组近邻值； k 指近邻值的个数； d_i 是当前异常数据状态向量与历史数据库中第 i 组状态向量数据之间的欧式距离； d_i 是第 i 个近邻值的权重。

2 实验验证

2.1 评价指标

本文采用平均相对误差为主要评价指标，均方根误差、相关系数评价指标作为辅助对交通流异常数据修复的有效性进行评价。平均相对误差 (MAPE) 表示修复值与真实值之间相对误差的均值，其值越小表示修复效果越好。均方根误差 (RMSE) 表示修复值和真实值之间的偏差，可以反映其修复能力，误差值低表示效果好。计算公式分别为：

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{v}_i - v_i}{v_i} \right| \times 100\% \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{v}_i - v_i)^2}{n}} \quad (8)$$

式中， v_i 是真实值， \hat{v}_i 是修复后的数值， n 是异常数据的个数。

相关系数 r 表示修复值和真实值之间的相似程度，其值大小范围为 $[-1, 1]$ 。其绝对值大小代表了数据之间的相关程度，负号表示数据关系为负相关，其绝对值越接近 1 表示相关性越强。计算公式如下：

$$r = \frac{\sum_{i=1}^n (v_i - mv_i)(\hat{v}_i - m\hat{v}_i)}{(\sum_{i=1}^n (v_i - mv_i)^2 \sum_{i=1}^n (\hat{v}_i - m\hat{v}_i)^2)^{\frac{1}{2}}} \quad (9)$$

式中， mv_i 表示数据 v_i 的均值， $m\hat{v}_i$ 表示数据 \hat{v}_i 的均值。

2.2 数据来源与处理

本文选取山东某高速公路监测点获取的实际交通流速度数据作为实验样本数据。为了更好地反映交通流量的运行情况,把交通速度数据合成为 5 分钟时间间隔的速度数据。选取 60 天正常交通流速度数据作为实验样本数据,其中 90% 作为历史数据库样本,10% 作为测试数据样本。部分交通流速度数据示例如表 1 所示。

表 1 部分交通流速度数据示例

交通流速度(km/h)					
时间	10月2号	10月3号	10月4号	10月5号	10月6号
7:00	87.17	101.55	105.87	98.90	75.50
7:05	97.11	124.31	131.05	91.87	88.86
7:10	93.06	115.74	87.21	98.62	106.79
7:15	110.99	111.77	100.37	91.60	111.40
7:20	113.53	113.95	101.26	91.46	84.03
7:25	127.90	109.58	128.11	91.32	102.92
7:30	115.49	118.67	109.77	102.39	105.60
7:35	105.08	111.39	114.28	61.25	106.78
7:40	110.09	94.05	105.71	106.88	102.24
7:45	112.86	102.12	87.84	125.73	93.00
7:50	99.21	96.22	107.19	94.08	100.41
7:55	104.19	110.28	96.00	131.58	100.07
8:00	116.26	113.56	108.71	104.98	115.18

2.3 结果比较与讨论

2.3.1 不同 k 值的结果比较

为了检验改进后的 KNN 数据修复模型的整体修复效果,用实验数据分析了 3 种 KNN 模型进行数据修复的平均相对误差和近邻值 k 的关系,并与常用的数据修复方法移动平均法相对比,结果如图 4 所示。

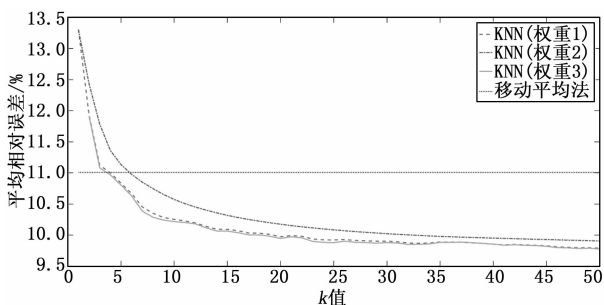


图 4 平均相对误差与 k 值的关系

从图 4 中可以看出:

1) 当 k 值大于 6 时,带权重 2 的 KNN 模型的修复效果优于移动平均法,当 k 值大于 4 时,带权重 1、3 的 KNN 模型修复效果优于移动平均法。由此可以得出,相比较移动平均法,KNN 修复模型的异常数据修复效果明显较好,说明 KNN 模型适用于数据修复,采取 k 近邻的思想进行数据修复的方法可行。

2) 随着 k 值的增大,3 种 KNN 数据修复模型的平均相对误差不断降低,并逐渐趋于平缓,权重 3 的 KNN 修复模

型所得修复结果的平均相对误差值比权重 1 和权重 2 修复模型的误差值小,说明改进的 KNN 数据修复模型的整体修复效果优于另两种权重 KNN 模型的整体修复效果。

2.3.2 不同模型结果比较

将平均相对误差、均方根误差、相关系数 3 种指标作为修复算法的误差评估方式对实验结果进行全面评价。将 3 种 KNN 数据修复模型和传统移动平均法 4 种方法的修复结果进行比较,见图 5、图 6 和表 2。

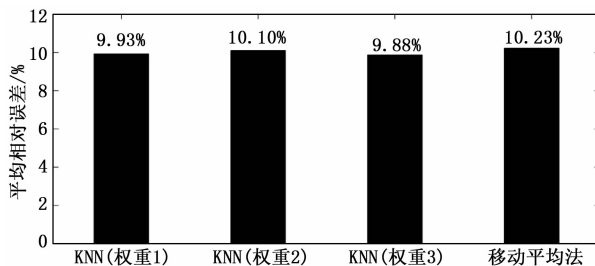


图 5 平均相对误差

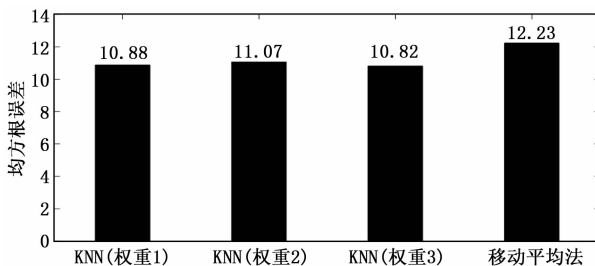


图 6 均方根误差

表 2 相关系数 r

修复方法		相关系数 r
KNN 修复模型	权重 1	0.72
	权重 2	0.71
	权重 3	0.73
移动平均法		0.66

图 5 显示了 4 种修复方法实验结果的平均相对误差 (MAPE) 结果比较。平均相对误差反映了交通流量数据修复值偏离真实值的程度。其值越小表示修复值与真实值的偏离程度越小,说明两者越相近。

从图 5 中可以看出,4 种修复方法实验结果的平均相对误差值大小依次为 9.93%、10.1%、9.88%、10.23%。由误差值大小可以看出,KNN 修复模型的误差值大小明显小于移动平均法,权重 3 修复模型的误差值要小于另两种权重修复模型。

图 6 显示了 4 种修复方法实验结果的均方根误差 (RMSE) 结果比较。表 1 显示了采用 4 种方法进行数据修复实验的修复值与真实值之间的相关系数 r 。相关系数表示修复数据与实际数据之间的相似性,其值越接近 1 表示修复值与真实值的相关度越高,修复效果越好。

从图 6 可以看出,4 种修复方法实验结果的均方根误差值大小依次为 10.88、11.07、10.82、12.23。从表 1 可以

得知, 在 KNN 修复模型中, 权重 3 的 KNN 修复模型的修复值与真实值的相关系数最大, 为 0.73, 权重 1 和权重 2 的 KNN 修复模型次之, 分别为 0.72 和 0.71, 移动平均法的相关系数最小, 为 0.66。

由此可以得出, 3 种 KNN 数据修复模型的修复效果明显优于移动平均法的修复效果, 权重 3 的 KNN 修复模型的修复精度明显高于权重 2 的修复模型, 略高于权重 1 的修复模型, 其平均相对误差大小为 9.88%。由此可以说明采用改进后的 KNN 异常数据修复模型进行数据修复取得了较好的修复效果。

为了更好的体现改进 KNN 修复模型的优势, 本文还分析了各个方法实验结果的相对误差在不同误差范围内所占的比例。图 7 表示采用 4 种修复方法实验结果的修复值与真实值的相对误差在 $[0, 20\%)$ 、 $[20\%, +\infty)$ 内的数据个数所占比例。可以看出, 采用 KNN 修复模型的相对误差小于 20% 的占比高于移动平均法, 大于 20% 的占比低于移动平均法。在 KNN 修复模型中, 权重 3 的修复模型, 其相对误差值在 $[0, 20\%)$ 范围内的占比最大。这说明改进后 KNN 修复模型的修复效果能够满足数据修复精度需求。

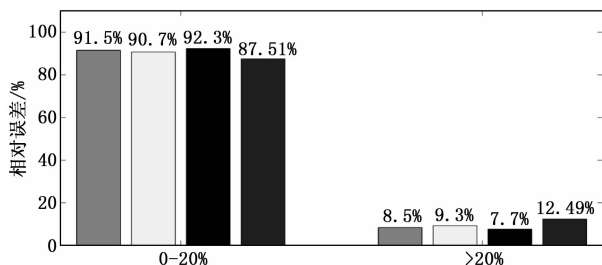


图 7 相对误差占比

3 结论

针对目前交通流数据异常和数据质量不高的问题, 将 k 近邻模型引用到数据修复中。对 KNN 基础模型相关要素进行改进和优化, 建立了一种适用于交通流数据修复的 KNN 异常数据修复模型。通过实际数据验证, 可得出以下结论:

1) 提出基于改进 KNN 算法的异常数据修复模型, 该修复模型与传统的移动平均修复方法相比, 具有更优的修复精度, 其平均相对误差值为 9.88%, 能够满足基本的异常数据修复精度需求。

2) 针对 KNN 算法模型, 进行 k 值和状态向量的优选, 提出一种距离占比的近邻值权重选取方式, 进而实现对基础模型的改进。结果表明, 采用距离占比的近邻值权重方式的 KNN 数据修复模型能够有效提高数据修复精度, 具有更好的数据修复效果。

本文所提出的数据修复方法实施性较强, 可有效提升交通流数据质量, 其修复思想对交通数据在实际应用方面具有理论指导意义。

参考文献:

[1] Chen H, Bell M. Instrumented city database analysts using multi-agents [J]. Transportation Research Part C Emerging

Technologies, 2002, 10 (5-6): 419-432.

- [2] Duan Y, Lv Y, Liu Y L, et al. An efficient realization of deep learning for traffic data imputation [J]. Transportation Research Part C, 2016, 72: 168-181.
- [3] 陆化普, 孙智源, 屈闻聪. 基于时空模型的交通流故障数据修正方法 [J]. 交通运输工程学报, 2015, 15 (6): 92-100, 117.
- [4] 姜桂艳, 江龙晖, 张晓东, 等. 动态交通数据故障识别与修复方法 [J]. 交通运输工程学报, 2004 (1): 121-125.
- [5] 胡伟超. 实时交通流数据缺失修复方法对比研究 [A]. 中国智能交通协会. 第八届中国智能交通年会论文集 [C]. 中国智能交通协会, 2013: 5.
- [6] Gold D L, Turner S, Gajewski B J. Imputing missing values in its data archives for intervals under 5 minutes [A]. Transportation Research Board 80th Annual Meeting [C]. 2001.
- [7] 王薇, 程泽阳, 刘梦依, 等. 基于时空相关性的交通流故障数据修复方法 [J]. 浙江大学学报 (工学版), 2017, 51 (9): 1727-1734.
- [8] 郭敏, 蓝金辉, 李娟娟, 等. 基于灰色残差 GM (1, N) 模型的交通流数据恢复算法 [J]. 交通运输系统工程与信息, 2012, 12 (1): 42-47.
- [9] Troyanskaya O, Cantor M, Sherlock G, et al. Missing value estimation methods for DNA microarrays [J]. Bioinformatics, 2001, 17 (6): 520-525.
- [10] Gu P, Yan X. Neural network approach to the reconstruction of freeform surfaces for reverse engineering [J]. 1995, 27 (1): 59-64.
- [11] Smith B, Scherer W, Conklin J. Exploring Imputation Techniques for Missing Data in Transportation Management Systems [J]. Transportation Research Record Journal of the Transportation Research Board, 2003, 1836 (1): 132-142.
- [12] Aydilek I B, Arslan A. A novel hybrid approach to estimating missing values in databases using K-nearest neighbors and neural networks [J]. Innovative Computing Information & Control, 2012, 8 (7): 4705-4717.
- [13] 袁媛, 邵春福, 林秋映, 等. 基于 RBF 神经网络的交通流数据修复研究 [J]. 交通运输研究, 2016, 2 (5): 46-52.
- [14] Davis G A, Nihan N L. Nonparametric Regression and Short-Term Freeway Traffic Forecasting [J]. Transportation Engineering, 2015, 117 (2): 178-188.
- [15] 宫晓燕, 汤淑明. 基于非参数回归的短时交通流量预测与事件检测综合算法 [J]. 中国公路学报, 2003, 16 (1): 82-86.
- [16] 张晓利, 贺国光, 陆化普. 基于 K 邻域非参数回归短时交通预测方法 [J]. 系统工程学报, 2009, 24 (2): 178-183.
- [17] 于滨, 邬珊华, 王明华, 等. K 近邻短时交通流预测模型 [J]. 交通运输工程学报, 2012, 12 (2): 105-111.
- [18] 谢海红, 戴许昊, 齐远. 短时交通流预测的改进 K 近邻算法 [J]. 交通运输工程学报, 2014, 14 (3): 87-94.
- [19] Habtemichael F G, Cetin M. Short-term traffic flow rate forecasting based on identifying similar traffic patterns [J]. Transportation Research Part C, 2016, 66: 61-78.
- [20] Liu H, Duan Z, Shi X. Correlation Analysis and Data Repair of Loop Data in Urban Expressway based on Co-integration Theory [J]. Procedia - Social and Behavioral Sciences, 2013, 96: 798-806.