

基于高斯过程回归的短时交通量不确定性预测

廖于¹, 廖梦媛¹, 刘雅澜¹, 刘佳嘉², 杨佳麒¹, 康杰¹

¹重庆交通大学信息科学与工程学院, 重庆

²重庆交通大学交通运输学院, 重庆

收稿日期: 2023年4月10日; 录用日期: 2023年5月24日; 发布日期: 2023年5月31日

摘要

智能交通预测对解决交通难题具有重要意义。针对现有模型预测不精准的问题, 本文提出一种全新的短期交通量预测模型——高斯回归模型(GPR), 这是基于贝叶斯理论和统计学习理论发展起来的机器学习算法。首先对数据进行归一化处理, 将数据划分为训练集和测试集, 建立标准高斯回归方程, 利用贝叶斯后验概率公式, 建立高斯回归预测模型, 并采用无效覆盖率(KP)和区间宽度(Width)两个评价指标, 将GPR与GARCH、BOOTSTRAP模型做对比, 结果显示本模型表现出的性能最优。

关键词

交通量预测, 高斯回归, 机器学习, 贝叶斯后概率公式

Short-Term Traffic Volume Uncertainty Prediction Based on Gaussian Process Regression

Yu Liao¹, Mengyuan Liao¹, Yalan Liu¹, Jiajia Liu², Jiaqi Yang¹, Jie Kang¹

¹School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing

²College of Traffic and Transportation, Chongqing Jiaotong University, Chongqing

Received: Apr. 10th, 2023; accepted: May 24th, 2023; published: May 31st, 2023

Abstract

Intelligent traffic prediction plays an important role in solving traffic problems. To solve the problem of inaccurate prediction of existing models, this paper proposes a new short-term traffic volume prediction model—Gaussian regression model (GPR), which is a machine learning algorithm developed based on Bayesian theory and statistical learning theory. Firstly, the data was

文章引用: 廖于, 廖梦媛, 刘雅澜, 刘佳嘉, 杨佳麒, 康杰. 基于高斯过程回归的短时交通量不确定性预测[J]. 交通技术, 2023, 12(3): 220-227. DOI: 10.12677/ojtt.2023.123025

normalized and divided into training set and test set, and the standard Gaussian regression equation was established. The Gaussian regression prediction model was established by using Bayesian posterior probability formula. Two evaluation indexes of invalid coverage (KP) and interval Width were used to compare GPR with GARCH and BOOTSTRAP models. The results show that the performance of this model is optimal.

Keywords

Traffic Volume Prediction, Gaussian Regression, Machine Learning, Bayesian Post Probability Formula

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

目前, 交通拥堵和效率低下已成为一个全球性问题。发达国家和发展中国家城市人口急剧增加所造成的最直接问题就是车辆数量的增加和交通堵塞。提高交通效率是改善日常交通状况的重要方法, 当下大数据的兴起也为提高交通效率提供了新的解决方案。各个城市每天都会产生大量的交通数据, 随着大数据处理的快速发展, 这种存储的数据信息开始发挥作用。通过对大数据中的数据信息进行分析和处理, 可以提出解决交通问题的新思路。

当前, 全球对智能交通系统[1]的投资正在增加。智能交通系统有许多环节, 包括交通量预测, 在智能交通控制和动态控制中起着重要的作用。借助准确的预测结果, 通勤者可以选择合适的交通方式、出行路线和出发时间, 实现诱导前提。同时, 实时、准确的路网短期交通流量预测有助于更好地分析路网交通状况[2], 在道路网络的交通规划和交通管理优化中发挥着重要作用。在构建智慧城市中, 实时预测交通量是亟待解决的问题。

现有的短期交通量预测模型[3]-[16], 如神经网络方法[7], 具有复杂的误差面。如果在计算过程中不小心, 神经网络会陷入局部极小点, 容易造成评估结果出现偏差。还有一些 AR 模型[6], 其中相位会影响频谱峰值的位置, 噪声会降低频谱估计等。此外, 交通量预测也存在数据缺失、冗余等问题, 也会受上下游路段流量影响。交通量数据受各方面因素的影响, 具有稳定性差、随机性强、高度非线性等特点, 使得交通量预测非常困难。

综上所述, 现有模型都存在着不容忽视的缺点, 结合交通量的复杂性和易受其他因素影响的特点, 可以得出结论, 目前用于交通流预测的模型不足以满足本文的交通量预测的真实需求。本团队提出了一种利用高斯过程回归模型(GPR) [15]的交通量预测方法。GPR 是一种全新的机器学习算法, 它是基于贝叶斯理论和统计学习理论发展起来的, 与 ARIMA [6]、支持向量积回归[3]。等方法相比, GPR 不仅能够预测未知量的期望值, 还能给出其分布状况; 与传统的区间预测方法相比, GPR 又具有机器学习算法泛化能力强等优点。此外, GPR 模型参数较少, 能自适应获取其超参数。因此 GPR 可以直接、方便地用于概率性预测[17]。GPR 有严格的统计学理论基础, 不仅能够处理低维数、小样本、非线性等复杂回归问题, 而且也适用于大样本和高维情形的扩展算法, 还具有参数自适应获取和预测结果具有概率意义等优点。并且 GPR 模型在其他领域已经取得了成功应用[18]。综上所述, 高斯过程回归模型满足了本文对交通量预测的实际需求。

2. 基于高斯回归过程的短时交通量预测

2.1. 高斯回归过程[19]

高斯过程是随机过程的一种, 适用于处理小样本、随机性强及含多维复杂因素的饱和负荷预测问题 [18] [20] [21] [22] 对于训练集 $D = \{(X, y) | X \in R^{n \times d}, y \in R^n\}$, 其中 $X = [x_1, x_2, x_3, \dots, x_n]^T$ 为一组训练输入变量, $Y = [y_1, y_2, y_3, \dots, y_n]^T$ 为目标输出向量。输入变量的随机过程状态集合 $f(X) = (f(x_i))$ 服从 n 维联合高斯分布, 因此 f 属于高斯过程, 其概率函数用 GP 表示, 可通过均值函数 $\rho(X)$ 及协方差函数矩阵 $K(X, X')$ 确定 [23], 即

$$f(X) \sim GP(\rho(X), K(X, X')) \quad (1)$$

高斯过程回归模型将输入变量 X 与目标输出 y 之间的关系视作高斯过程 f ; 把独立的白噪声 ε 考虑到 f , 则可建立标准高斯过程回归模型:

$$y = f(X) + \varepsilon \quad (2)$$

式中: ε 为白噪声, 服从高斯分布, 即满足 $\varepsilon \sim N(0, \sigma_n^2 I)$, I 为单位阵, σ_n^2 为方差。由于白噪声具有独立性, 因此 y 同样属于高斯过程, 即

$$y \sim GP(\rho(X), K(X, X') + \sigma_n^2 I) \quad (3)$$

根据贝叶斯原理, 在给定的训练集 D 内, 建立 y 的先验分布:

$$y \sim N(0, K(X, X') + \sigma_n^2 I) \quad (4)$$

则对测试样本 $\{(x^*, y^*) | x^* \in R^d, y^* \in R\}$, 根据高斯过程的性质, 训练样本的目标输出 y 与测试样本输出 y^* 服从联合高斯分布:

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim N \left(0, \begin{bmatrix} K(X, X') + \sigma_n^2 I & K(X, x^*) \\ K(x^*, X) & K(x^*, x^*) \end{bmatrix} \right) \quad (5)$$

式中: $K(X, X) = (k(x_i, x_j))$ 为训练输入变量 X 的 $n \times n$ 阶协方差函数矩阵, $k(x_i, x_j)$ 为核函数;

$K(X, x^*) = (k(x_i, x^*))$ 为训练输入变量 X 与测试输入变量 x^* 的 $n \times 1$ 阶协方差函数矩阵;

$K(x^*, x^*) = (k(x^*, x^*))$ 为测试输入变量 x^* 自身的协方差。

利用贝叶斯后验概率公式, 在给定测试输入变量 x^* 与训练集 D^* 的条件下, 对应的输出 y^* 满足:

$$y^* | x^*, D^* \sim N(\bar{y}^*, \text{cov}(y^*)) \quad (6)$$

式中: \bar{y}^* 和 $\text{cov}(y^*)$ 分别是测试样本输出 y^* 的均值和方差。

则高斯过程回归预测模型为:

$$\bar{y}^* = K(x^*, X) (K(X, X) + \sigma_n^2 I)^{-1} y \quad (7)$$

$$\text{cov}(y^*) = K(x^*, x^*) - K(x^*, X) (K(X, X) + \sigma_n^2 I)^{-1} K(X, x^*) \quad (8)$$

利用高斯分布的“ 3σ 原理”, 测试样本输出 y^* 预测值 99.73% 的置信区间为:

$$\left[\bar{y}^* - 3\sqrt{\text{cov}(y^*)}, \bar{y}^* + 3\sqrt{\text{cov}(y^*)} \right] \quad (9)$$

2.2. 分析步骤

用高斯回归过程精准预测短时交通量。如图 1 所示。

- 1) 收集若干交通量的历史时间序列数据，得到 $x_i (i=1,2,\dots)$ 。
- 2) 当交通量数值变化差异较大时，会对高斯过程学习造成不利影响，因此需要对交通量进行归一化处理，具体为：

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (10)$$

式中： $\max(x)$ 和 $\min(x)$ 为最大值和最小值； x 为原始值。

- 3) 交通量数据划分为训练集和测试集，建立标准高斯回归过程模型，根据贝叶斯原理，建立 y 的先验分布。

- 4) 对高斯过程回归模型进行参数选择和调整之后，将训练所得模型在测试样本上进行测试，通过评价指标可知，高斯过程回归模型对数据的拟合效果较好[24]。

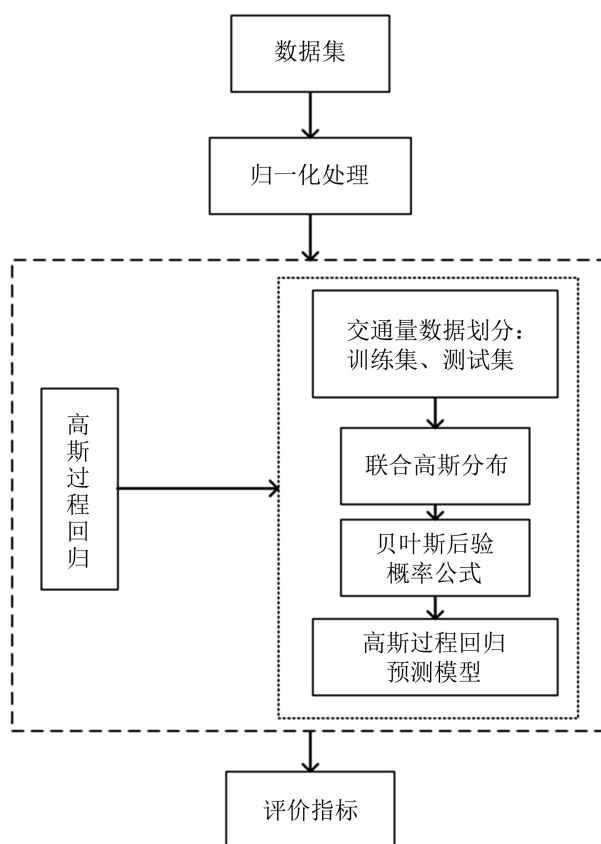


Figure 1. Flowchart of the analysis procedure

图 1. 分析步骤流程图

3. 案例应用与分析

3.1. 数据说明

为了测试 GPR 模型的短时交通量预测性能，数据采用重庆市某主干道 2019-10-11 00:00:00 至

2019-10-31 20:00:00 的每 5 分钟的交通流量，得到 6000 个样本，如图 2 所示。前 3000 个样本作为训练集，其余样本作为测试集。

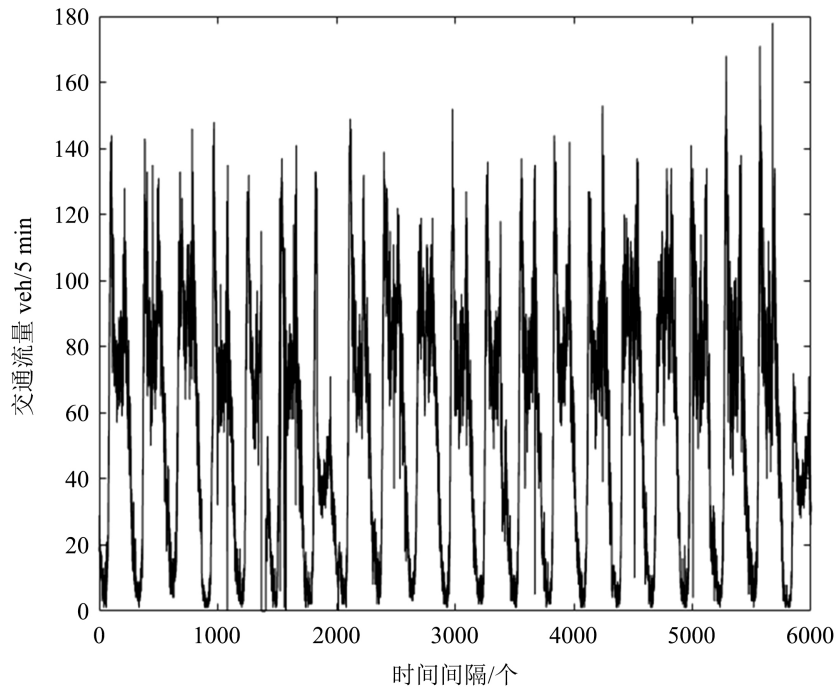


Figure 2. Traffic flow
图 2. 交通流量

为了更加方便地使用数据，消除奇异样本点对整体数据的不良影响，将采集到的数据进行归一化处理。

3.2. 评价指标

针对短时交通量不确定性区间的特点，本文采用无效覆盖率(KP)和区间宽度(Width)两个评价指标，它们定义为：

$$KP = n^{-1} \sum I(\circ) \times 100\% \quad (11)$$

式中：

$$I(\circ) = \begin{cases} 0, & x_i^{low} < x_i < x_i^{high} \\ 1, & x_i < x_i^{low} \text{ or } x_i > x_i^{high}, i = 1, \dots, n \end{cases} \quad (12)$$

$$Width = \frac{1}{n} \sum_{i=1}^n \sqrt{(x_i^{low} - x_i)^2 + (x_i^{high} - x_i)^2} \quad (13)$$

式中各表达式含义与 KP 相同[25]。

3.3. 预测结果分析

将交通量数据进行划分后，对数据进行归一化处理。在 MATLAB 中，再对已经归一化处理后的数据集进行高斯回归过程处理。预测结果如图 3 所示：

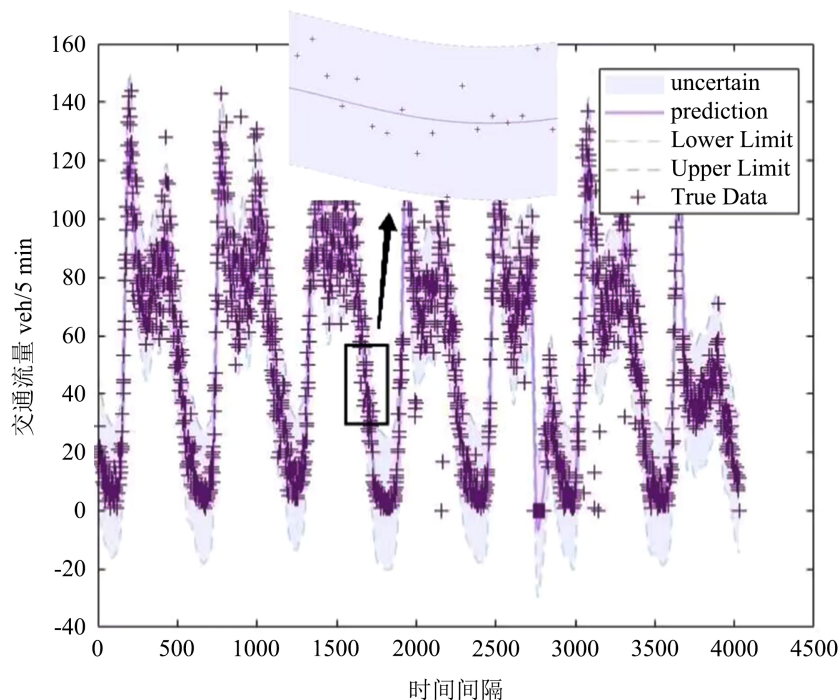


Figure 3. Model prediction

图 3. 模型预测

本文使用 GARCH [7]、BOOTSTRAP 模型[25]与本文所提模型进行预测结果评价指标的对比, 各模型的评价指标如表 1 所示:

Table 1. Evaluation indicators of different models

表 1. 不同模型的评价指标

预测模型	KP	Width
GPR	5.40%	216.54
GARCH	6.00%	197.95
BOOTSTRAP	5.64%	210.20

由表 1 可知, 本文模型相较于 GARCH, KP 和 Width 分别提高了 0.6%、18.59; 相较于 BOOTSTRAP, KP 和 Width 分别提高了 0.24%、6.34。结果表明, 本文模型表现出的性能最优。

4. 结论

高斯过程回归模型(GPR)具有机器学习算法泛化能力强、模型参数少、超参数自适应采集、统计学理论基础严格等优点, 可以直接方便地用于概率性预测。在此基础上, 本文建立了高斯过程回归模型作为道路交通量状态动态分析和预测的方法。并通过将高斯过程回归模型与 GARCH、BOOTSTRAP 模型进行对比, 发现高斯过程回归模型表现出的性能最优, 且对数据有很好的拟合效果, 满足了我们一开始希望通过高斯过程回归模型预测交通量、减少交通拥堵等问题的设想。综上, 高斯过程回归模型满足了本文对交通量预测的实际需求。

基金项目

重庆交通大学大学生创新创业训练计划项目《城市道路短时交通流预测研究——以重庆市某主干道为例》(S202210618027)资助。重庆交通大学大学生创新创业训练计划项目《一种新型混合模式在短时交通流预测中的应用研究》(S202210618039)资助。

参考文献

- [1] Lv, Y., Duan, Y., Kang, W., Li, Z. and Wang, F.Y. (2015) Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, **16**, 865-873.
- [2] 焦琴琴. 基于深度学习的路网短时交通流预测[D]: [硕士学位论文]. 西安: 长安大学, 2016.
- [3] 张玺君, 余光杰, 崔勇, 尚继洋. 基于聚类算法和图神经网络的短时交通量预测[J/OL]. 吉林大学学报(工学版): 1-9. <https://kns.cnki.net/kcms/detail/22.1341.t.20221104.1545.008.html>, 2022-11-07.
- [4] 傅贵, 韩国强, 遼峰, 等. 基于支持向量机回归的短时交通量预测模型[J]. 华南理工大学学报(自然科学版), 2013, 41(9): 71-76.
- [5] 吕磊. 短时交通量预测与线路推荐研究[D]: [硕士学位论文]. 济南: 山东大学, 2016.
- [6] 朴春俊, 蔡满军, 吴忠强. 一种新的自回归——滑动平均模型[J]. 信息技术, 2001(8): 1-3.
- [7] 谭满春, 冯萃斌, 徐建闽. 基于 ARIMA 与人工神经网络组合模型的交通流预测[J]. 中国公路学报, 2007, 20(4): 118-121.
- [8] 杨明极, 张贵山. 基于栈式自动编码器的语音质量评价方法[J]. 小型微型计算机系统, 2018, 39(10): 2134-2137.
- [9] 姜宗辰, 马毅, 江涛, 等. 基于深度置信网络(DBN)的赤潮高光谱遥感提取研究[J]. 海洋技术学报, 2019, 38(2): 1-7.
- [10] Han, L. and Huang, Y.S. (2020) Short-Term Traffic Flow Prediction of Road Network Based on Deep Learning. *IET Intelligent Transport Systems*, **14**, 495-503. <https://doi.org/10.1049/iet-its.2019.0133>
- [11] Li, J., Guo, F., Sivakumar, A., Dong, Y.J. and Krishnan, R. (2021) Transferability Improvement in Short-Term Traffic Prediction Using Stacked LSTM Network. *Transportation Research Part C: Emerging Technologies*, **124**, Article ID: 102977. <https://doi.org/10.1016/j.trc.2021.102977>
- [12] Wang, Q.M., Fan, A.W. and Shi, H.S. (2017) Network Traffic Prediction Based on Improved Support Vector Machine. *International Journal of System Assurance Engineering and Management*, **8**, 1976-1980. <https://doi.org/10.1007/s13198-016-0412-8>
- [13] Wu, C.H., Wei, C.C., Su, D.C., Chang, M.H. and Ho, J.M. (2003) Travel Time Prediction with Support Vector Regression. *Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems*, Shanghai, 12-15 October 2003, 1438-1442.
- [14] Zheng, Z., Chen, W.H., Wu, X.M., Chen, P.C.Y. and Liu, J.M. (2017) LSTM Network: A Deep Learning Approach for Short-Term Traffic Forecast. *IET Intelligent Transport Systems*, **11**, 68-75. <https://doi.org/10.1049/iet-its.2016.0208>
- [15] Guo, Z.Q., Zhao, X., Chen, Y.X., Wu, W. and Yang, J. (2019) Short-Term Passenger Flow Forecast of Urban Rail Transit Based on GPR and KRR. *IET Intelligent Transport Systems*, **13**, 1374-1382. <https://doi.org/10.1049/iet-its.2018.5530>
- [16] Zhou, B., Ma, X.J., Luo, Y.H. and Yang, D.S. (2019) Wind Power Prediction Based on LSTM Networks and Nonparametric Kernel Density Estimation. *IEEE Access*, **7**, 165279-165292. <https://doi.org/10.1109/ACCESS.2019.2952555>
- [17] 宗文婷, 卫志农, 孙国强, 等. 基于改进高斯过程回归模型的短期负荷区间预测[J]. 电力系统及其自动化学报, 2017, 29(8): 22-28.
- [18] 黄亚, 易灵, 肖伟华, 等. 基于高斯过程回归模型的径流短期预测研究[J]. 水力发电, 2020, 46(12): 9-12.
- [19] 彭虹桥, 顾洁, 胡玉, 等. 基于混沌粒子群——高斯过程回归的饱和负荷概率预测模型[J]. 电力系统自动化, 2017, 41(21): 25-32, 155.
- [20] 甘迪, 柯德平, 孙元章, 等. 基于集合经验模式分解和遗传——高斯过程回归的短期风速概率预测[J]. 电工技术学报, 2015, 30(11): 138-147.
- [21] Alamaniotis, M., Chatzidakis, S. and Tsoukalas, L.H. (2014) Monthly Load Forecasting Using Kernel Based Gaussian Process Regression. *MedPower 2014*, Athens, 2-5 November 2014, 1-8. <https://doi.org/10.1049/cp.2014.1693>
- [22] Yan, W., Qiu, H. and Xue, Y. (2009) Gaussian Process for Long-Term Time-Series Forecasting. 2009 *International*

Joint Conference on Neural Networks, Atlanta, 14-19 June 2009, 3420-3427.

<https://doi.org/10.1109/IJCNN.2009.5178729>

- [23] Rasmussen, C.E. and Williams, C.K.I. (2006) *Gaussian Processes for Machine Learning*. MIT Press, Massachusetts.
<https://doi.org/10.7551/mitpress/3206.001.0001>
- [24] 李振刚. 基于高斯过程回归的网络流量预测模型[J]. 计算机应用, 2014, 34(5): 1251-1254.
- [25] 闫晓楠. 基于改进 BOOTSTRAP 方法的短时交通流不确定性预测[D]: [硕士学位论文]. 南京: 东南大学, 2020.