

# 城市交通拥堵预测中大数据与数学建模的融合研究

刘世婕, 刘艳艳, 黄友霞, 李亚涵

北京邮电大学世纪学院, 北京

收稿日期: 2026年2月13日; 录用日期: 2026年3月6日; 发布日期: 2026年3月13日

## 摘要

本文摒弃大数据与数学建模的宏观综述视角, 以城市交通拥堵预测为具体应用场景, 系统梳理数学建模在该场景下的核心流程, 界定了大数据的三类数据类型及数据量、多样性、速度等关键特征, 并详细拆解了大数据分析在数学建模中的典型流程, 包括数据预处理、建模与分析、结果解释与可视化三大环节。同时, 本文探讨了分布式计算、卷积神经网络、遗传算法等主流算法与大数据分析方法的适配性, 列举了其在气象预报建模、癌症放疗精准治疗等典型场景中的应用实践。本文表明, 大数据分析通过数据降噪、特征优化等手段显著提升了数学模型的准确性与鲁棒性, 而数学建模则为海量复杂数据的分析提供了核心逻辑框架。未来, 随着实时数据处理技术的发展, 二者的深度融合将推动数学建模向“数据驱动的精准预测”方向迈进, 在复杂系统模拟、实时决策支持等领域发挥更大价值。

## 关键词

数学建模, 交通大数据分析, 算法优化, 数据处理, 交通拥堵预测

# Research on the Integration of Big Data and Mathematical Modeling in Urban Traffic Congestion Prediction

Shijie Liu, Yanyan Liu, Youxia Huang, Yahan Li

Century College, Beijing University of Posts and Telecommunications, Beijing

Received: February 13, 2026; accepted: March 6, 2026; published: March 13, 2026

## Abstract

This article moves away from the macro overview perspective of big data and mathematical modeling,

focusing instead on the specific application scenario of urban traffic congestion prediction. It systematically outlines the core process of mathematical modeling in this context, defines three types of big data along with key characteristics such as data volume, diversity, and velocity, and provides a detailed breakdown of the typical process of big data analysis in mathematical modeling, encompassing three major steps: data preprocessing, modeling and analysis, and result interpretation and visualization. Additionally, this article explores the adaptability of mainstream algorithms such as distributed computing, convolutional neural networks, and genetic algorithms to big data analysis methods, and lists their practical applications in typical scenarios like weather forecast modeling and precision radiotherapy for cancer treatment. This article demonstrates that big data analysis significantly enhances the accuracy and robustness of mathematical models through techniques such as data denoising and feature optimization, while mathematical modeling provides the core logical framework for analyzing massive and complex data. In the future, with the development of real-time data processing technology, the deep integration of the two will propel mathematical modeling towards the direction of "data-driven precise prediction", exerting greater value in areas such as complex system simulation and real-time decision support.

## Keywords

Mathematical Modeling, Traffic Big Data Analysis, Algorithm Optimization, Data Processing, Traffic Congestion Prediction

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

数学建模在理工科方面占有重要的地位，主要是通过构建可计算的数学模型解决实际问题。在整个过程中需要调查问题背景、定义问题参数、通过公式将实际问题转换为数学模型，进而完成建模、求解、验证、评价指标等流程。随着科技进步，AI在许多步骤上助益良多。例如，AI工具使得模型的可视化、算法的流程图等呈现得更加明了。这也就告诉我们在数学建模上提出问题、解决问题(例如通过概率论与数理统计、微分方程、算法优化等方法)成为数学建模的核心内容。

在人工智能以及大数据技术爆发的时代背景下，数据分析以及数学建模的交叉应用也需要与时俱进，不能完全依赖传统模式。数据分析的三大核心环节常出现于城市交通拥堵预测数学建模的实践环节中：数据预处理阶段，通过数据降噪、特征显著化提升数据质量；优化算法阶段，通过神经网络、随机森林、机器学习、模拟退火算法等优化模型(如卷积神经网络处理医学影像数据时，预测误差可控制在0.01~0.04) [1]；验证阶段，通过交叉验证、设计评价指标等方法提升算法的泛化能力(如人口流动时空特征建模中，结合时空大数据与SEIR模型改进版，其传播路径预测准确率较传统模型提升了30%以上[2])。

综上所述，大数据分析数学建模相辅相成。一方面，大数据分析通过数据预处理、降噪、特征优化等方式提升模型的准确性、泛化能力以及鲁棒性。另一方面，面对数据量与复杂程度与日俱增，数学建模在大数据分析的应用中同样提供至关重要的作用[3]。

## 2. 数学建模的特点以及在交通拥堵预测中的应用

数学建模在计算领域应用广泛。简而言之，数学建模是一种运用特定方式将复杂且棘手的问题转化为可有效处理的模型构建过程。该过程基于严密的数学定律与规范程序，因而具备高度的归纳性[4]。

随着数学与程序日益紧密结合, 数学建模已成为跨语言交流的有效方式。数学建模能识别并提炼出信息中的核心数学问题, 对其进行提炼概括, 进而完成分析与总结。在实际应用中, 数学建模通过提炼规则、提出假设、构建结构模型, 并利用该模型进行计算与分析, 最终实现预设的分析目标。

数学建模的这些特点和优势能够很好地应用于城市交通拥堵预测模型中, 为其提供数据支持以及优化参考, 让程序更加完善, 使用户体验得以提升[5]。然而, 在实际应用中, 数学建模对于计算机程序设计的优化效果仍需持续探索实践。

### 3. 交通大数据分析及结构与特点

大数据分析[6]涵盖了诸多用于分析数据、提取信息以及从带有复杂模式的大规模数据集中获取见解的技术, 而传统的数据处理方式很难对其进行处理。大样本赋予其更强的统计效力, 但面对高维数据与复杂模型时, 分析过程易受干扰, 从而导致错误发现率上升。

#### 3.1. 交通大数据可划分为以下三类

**结构化数据:** 结构化数据有着清晰明确的格式, 能够十分便捷地以表格形式存储于像 MySQL、Oracle 这类关系数据库当中。

**半结构化数据:** 半结构化数据具备一定的结构, 但无法按照表格格式记录在关系数据库里。像 XML 文件、JSON 文档、电子邮件消息等都属于半结构化数据的范畴。

**非结构化数据:** 非结构化数据是没有结构的, 无法以表格形式存放在关系数据库中。例如视频、音频、文本、机器生成的数据等都属于非结构化数据。

#### 3.2. 交通大数据的特点

交通大数据具备诸多特征, 具体如下:

**数据量:** 指每天都在快速增长的数据规模。人类的活动、各类技术的运用、人们在社交媒体上的互动等, 都会产生海量的数据。

**多样性:** 大数据来源的丰富性使得数据类型多样化。这些数据可能是结构化的、半结构化的或者非结构化的。单个应用常生成或收集多类数据, 需关联整合方可提炼知识。

**速度:** 指来自不同交通监测设备、交通出行平台持续不断的数据流, 并且数据存储库也会以同样的速率不断更新数据。要及时采集数据流以供处理已成为一项挑战。

**真实性:** 指交通数据存在不可靠的情况。具体表现为数据不一致、不完整, 或包含大量噪声(如交通监测设备的误差、出行数据中的异常标注、缩写、拼写错误等), 导致其质量与准确性难以有效控制。

**可变性:** 对于交通数据来说, 研究背景、数据特征等并非恒定, 交通数据的应用场景和研究方向的变化, 会对其产生显著的影响。

**可视化:** 通过表格、图形等方式呈现数据或结果图。

### 4. 交通大数据分析在拥堵预测数学建模中的典型流程

#### 4.1. 数据预处理: 夯实拥堵预测建模基础

**数据清洗:** 在交通拥堵预测数学建模的过程中, 数据质量至关重要。数据清洗则是保障的第一步骤, 主要是处理缺失值、异常值等噪声问题。对于缺失值来说, 常用取均值填充和多重插补等方法进行补充缺失值。

异常值的处理同样不容小觑, 例如可依照  $3\sigma$  原则来剔除异常值:  $3\sigma$  原则是依托正态分布的特性而

确立的，通常，均值加减 3 倍标准差之外的数据点被视为异常值。将其去除可避免对模型造成偏差影响。基于更合理的数据进行训练与分析，进而提升其可靠性与准确性。噪声会干扰模型捕捉数据本质特征，这时可运用滤波算法来去除。滤波算法有着不同的原理：有的侧重于将数据曲线进行平滑处理，有的则基于频率分析等手段去除特定频段的干扰信号。从而使数据更干净、平滑，更能体现出真实的内在规律。

**特征工程：**特征工程在数据预处理环节占据着极为关键的地位，其核心在于从原始数据中挖掘出更有助于模型学习与分析的有效信息，主要涵盖特征提取、特征变换以及特征选择这几个方面。

**特征提取：**面对复杂的原始交通数据，例如海量交通监控图像，直接将其作为模型的输入，往往难以取得理想的效果。此时，可以借助卷积神经网络(CNN)从中提取关键特征。以交通图像数据为例，CNN能够提取出诸如车流量密度、车辆行驶轨迹、道路拥堵区域边缘特征等，这些经过提炼的特征更能体现交通运行的本质内容，对后续的拥堵分类、趋势预测等任务大有裨益，并且还能大幅削减数据的维度，提高模型的运行效率与准确性。

**特征变换：**不同特征的量纲往往存在差异，这很可能会给模型的训练与评估带来不利影响。例如，一个特征的取值范围处于 0 到 100 之间，而另一个特征的取值范围可能是 0 到 10,000，为消除这种量纲方面的影响，需要进行特征变换操作。标准化是依据数据的均值和标准差来实施线性变换的，把数据转变为均值为 0，标准差为 1 的分布形态，比较适用于多数符合正态分布的数据情况。归一化则是把数据映射到 0 到 1 的区间内，通过原始数据与最小值、最大值之间的差值关系来完成变换，常用于数据取值范围差异较大且对数据区间范围有明确要求的场景之中。借助特征变换，各个特征能够处于相对公平的量纲水平，这样一来，模型就能更聚焦于特征之间的内在联系，而不会受量纲差异的误导。

**特征选择：**原始交通数据中的冗余变量既会增加模型复杂度，也可能引发过拟合等问题。为剔除这些冗余变量，可以采用方差过滤、递归特征消除(RFE)等方法。方差过滤是依照变量的方差大小进行筛选，方差较小意味着该变量的变化幅度不大，对模型的贡献程度可能较低，便可将其过滤掉。递归特征消除则是通过迭代构建模型并评估各个特征的重要性，逐步剔除次要特征，保留对模型最具价值的特征子集，以此降低模型的复杂度，提高模型的泛化能力以及可解释性。

## 4.2. 建模与分析：算法与数据的协同

### 4.2.1. 传统数学模型与交通数据分析结合：城市交通拥堵预测模型构建

**传统数学模型与数据分析结合：**传统数学模型在处理结构化数据方面有着独特的优势，当其与分析相互融合时，能够发挥出更为显著的作用[7]。

本文以城市道路路段交通拥堵预测为具体对象，完整展示模型构建的数学过程：

#### (1) 模型假设条件

- ① 研究对象为城市某单向通行的主干道路段，路段长度为  $L$ ，车道数为  $n$ ，道路通行能力稳定；
- ② 交通流为连续流，忽略车辆个体的行驶差异，以车流密度、平均车速、车流量作为核心分析指标；
- ③ 模型仅考虑路段上下游车流量、路段内车辆行驶速度、红绿灯配时、高峰时段系数等核心影响因素，暂不考虑交通事故、道路施工等突发因素；
- ④ 交通数据的采集时间粒度为 5 分钟，建模预测时间尺度为未来 15~60 分钟的路段拥堵等级。

#### (2) 核心方程建立

基于交通流理论，选取格林希尔治(Greenshields)交通流模型为基础，结合交通大数据分析得到的影响因子，构建拥堵预测方程。

- ① 基础交通流关系方程： $Q = K \times V$  其中， $Q$  为路段车流量， $K$  为车流密度， $V$  为路段平均车速。

② Greenshields 速度 - 密度线性关系方程:  $V = V_f \left(1 - \frac{K}{K_j}\right)$  其中,  $V_f$  为自由流车速, 由交通大数据统计得到;  $K_j$  为阻塞密度, 为路段最大容纳车流密度。

③ 结合高峰时段、红绿灯配时影响, 引入修正系数, 得到修正后速度 - 密度方程:  
 $V = V_f \left(1 - \frac{K}{K_j}\right) \times \alpha \times \beta$  其中,  $\alpha$  为高峰时段系数, 高峰时段  $\alpha = 0.6-0.8$ , 平峰时段  $\alpha = 1.0$ , 由交通大数据时序分析得到;  $\beta$  为红绿灯配时修正系数,  $\beta = \frac{t_g}{t_c}$ ,  $t_g$  为绿灯时长,  $t_c$  为红绿灯周期时长。

④ 拥堵等级判定方程: 通过交通大数据聚类分析, 将车流密度  $K$  作为拥堵等级核心判定指标, 划分 4 级拥堵标准, 构建判定方程:

$$Y = \begin{cases} 1, & K < K_1 \text{ 畅通} \\ 2, & K_1 \leq K < K_2 \text{ 轻度拥堵} \\ 3, & K_2 \leq K < K_3 \text{ 中度拥堵} \\ 4, & K \geq K_3 \text{ 重度拥堵} \end{cases}$$

其中,  $Y$  为拥堵等级,  $K_1$ 、 $K_2$ 、 $K_3$  为密度阈值, 由研究路段的交通大数据统计分析得到。

⑤ 拥堵预测核心方程: 结合时间序列分析, 引入前  $t-1$ 、 $t-2$  时刻的车流密度作为自回归因子, 构建基于 ARIMA 的拥堵预测密度方程, 进而通过等级判定方程得到拥堵预测结果:

$K_t = \varphi_1 K_{t-1} + \varphi_2 K_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$ , 其中,  $K_t$  为  $t$  时刻的预测车流密度,  $\varphi_1$ 、 $\varphi_2$  为自回归系数,  $\theta_1$  为移动平均系数,  $\varepsilon_t$  为随机误差项, 各系数由交通大数据拟合求解得到。

### (3) 求解算法步骤

① 数据输入: 将预处理后的交通大数据(历史车流密度、车速、车流量、高峰系数、红绿灯配时等)导入模型;

② 参数标定: 利用最小二乘法对 Greenshields 模型中的  $V_f$ 、 $K_j$  进行标定, 利用极大似然估计法对 ARIMA 模型中的  $\varphi_1$ 、 $\varphi_2$ 、 $\theta_1$  进行求解;

③ 密度预测: 将标定后的参数代入拥堵预测核心方程, 求解得到未来各时刻的车流密度  $K_t$ ;

④ 拥堵判定: 将  $K_t$  代入拥堵等级判定方程, 得到未来路段的拥堵等级  $Y_t$ ;

⑤ 结果输出: 输出未来 15~60 分钟的路段拥堵等级及对应的车流密度、车速预测值。

### (4) 模型稳定性数学证明

模型稳定性采用李雅普诺夫稳定性判据验证, 针对拥堵预测核心方程的离散时间系统:

$K_t = \varphi_1 K_{t-1} + \varphi_2 K_{t-2} + \theta_1 \varepsilon_{t-1}$ , 忽略随机误差项后得到齐次线性差分方程:  $K_t - \varphi_1 K_{t-1} - \varphi_2 K_{t-2} = 0$ , 其特征方程为:  $\lambda^2 - \varphi_1 \lambda - \varphi_2 = 0$ 。

求解特征根  $\lambda_1$ 、 $\lambda_2$ , 若所有特征根的模都小于 1, 即  $|\lambda_1| < 1$  且  $|\lambda_2| < 1$ , 则该离散时间系统是渐近稳定的, 模型的预测结果具有收敛性, 不会出现无界波动。

通过交通大数据拟合得到的自回归系数满足  $\varphi_1 + \varphi_2 < 1$ ,  $\varphi_2 - \varphi_1 < 1$ ,  $|\varphi_2| < 1$ , 可推导出特征根的模均小于 1, 因此该拥堵预测模型满足渐近稳定性, 预测结果具有可靠性。

## 4.2.2. 机器学习模型的交通数据分析应用

机器学习模型凭借其强大的学习能力以及对复杂数据的处理能力, 在数据分析领域的重要性日益凸显, 依据有无标签数据这一情况, 可将其划分为监督学习和无监督学习这两类应用场景[8]。

监督学习: 在存在标签交通数据的情况下, 监督学习中的分类和回归方法能够有效地挖掘交通数据

中的规律并进行预测。例如在拥堵等级分类任务里的随机森林算法，它通过构建多个决策树，并综合这些决策树的分类结果，以此提高预测的准确性和稳定性，在城市道路拥堵预测、交通出行需求预判等诸多场景中都有着广泛的应用。比如基于路段的历史交通数据，涵盖车流量、车速、道路占有率、高峰时段等多维度信息来构建拥堵等级预测模型，通过对已标注好拥堵等级的历史路段数据展开学习训练，随机森林模型就能精准地对路段未来的拥堵等级进行分类预测。

**无监督学习：**对于没有标签的交通数据，无监督学习中的聚类和降维方法能够助力我们探寻交通数据的内在分布结构。例如 K-Means 聚类算法，它按照数据点之间的距离远近将数据划分成不同的簇，使同簇内样本相似度高、不同簇间样本差异显著。在城市交通小区出行特征分析中，可以运用该算法依据交通小区的出行量、出行时间、出行方式等特征对小区进行聚类，把小区分成诸如高出行量通勤型、低出行量生活型等不同的群体，之后针对各个群体分别开展建模分析，制定出个性化的交通管控和疏导方案。

### 4.3. 结果解释与可视化

数据可视化是帮助我们掌握模型输入输出以及的数据特征等信息的最直观的方式。例如：热力图可视化：热力图通过颜色的深浅反映特征之间相关性的强弱，颜色越深相关性越强，这对后续的特征选择、模型解释等具有重要的参考价值；箱线图可以直观地展示变量的分布情况；散点图可以直观地展示数据之间的线性关系。

**模型可解释性技术：**随着机器学习和深度学习模型发展越来越复杂，理解其逻辑相对困难，因此，模型可解释性技术应运而生。

## 5. 算法类型与大数据分析方法

为应对交通数据的爆炸性增长，分布式计算与存储系统应运而生。分布式计算与存储系统能够高效地处理海量数据，它既能扩充储存量，又能确保数据的可靠性与可扩展性。例如在处理大量的图像数据、文本数据等时，分布式存储系统可以将这些数据合理地分配到不同的存储节点，从而避免单个存储设备出现容量不足的问题。与此同时，利用并行计算技术能够加快神经网络等复杂模型的训练进程。并行计算会把计算任务拆解成多个子任务，然后分配到多个计算节点上同步进行处理，这显著缩短了模型训练时间，使海量数据下的有效建模更为迅速。

另有 Hadoop 为研究对象的情况，深入分析其分布式文件系统(HDFS)和 MapReduce 模型在大数据处理中的性能优化问题[9]。HDFS 采用的是分布式存储架构，它会把大文件切割成多个数据块，分别存储在不同的节点上，并且借助副本机制来保障数据的安全性，即便某个节点发生故障，也能够从其他副本节点获取相应的数据。在数据分析与建模的过程中，涉及的算法类别丰富多样，涵盖了神经网络算法、传统数学算法等诸多类别，算法因场景而异，都发挥着独特的作用[10]。

随着大数据应用的拓展，从机器学习到深度学习等各类分析方法不断涌现，它们各具特点，针对不同的数据与分析需求发挥着重要作用。

**机器学习：**机器学习方法具备对大规模气候变化数据进行分析、解读以及预测的能力，它通过从大量的数据中学习规律和模式，能够迅速开展数据分析工作，并给出相对准确的预测结果，这对于改善清洁能源获取等方面有着重要的意义。例如，通过分析历史气候数据与清洁能源发电数据之间的关系，运用机器学习模型对未来的气候条件进行预测，进而合理安排清洁能源的生产与储备，提高能源利用效率。不过，机器学习方法依赖大规模训练数据以确保模型的准确性和泛化能力，且训练周期长、计算成本高，尤其在处理复杂数据时更为显著。

深度学习：深度学习作为一种模拟和预测的非线性方法，有着很强的表示能力，能够自动挖掘数据中的深层次特征和复杂模式，便于发现诸如气候模式等隐藏在数据中的规律，在气候数据挖掘和诊断中展现出良好的性能。例如，利用深度学习模型分析气候时序数据，可有效识别周期性规律、极端事件特征等。然而，该方法也存在明显局限：神经网络易出现过拟合，导致泛化能力下降；超参数调优耗时费力；且训练成本高，对 GPU 等硬件要求严苛，一定程度上限制了其在资源有限场景中的应用。

人工智能：人工智能涵盖了范围更广的技术和方法，它支持对数据进行模拟和决策，当与地球观测数据和气候模拟数据相结合时，能够发挥出更优的效果。例如，借助人工智能技术对卫星观测到的地球表面数据、大气数据等进行综合分析，模拟气候系统的变化，为气候预测、灾害预警等提供决策支持。但人工智能的实施成本较高，既需专业团队与充足算力，也面临高质量数据获取的难题，这在一定程度上制约了其在气候领域的广泛应用。

综上所述，不同的大数据分析方法在各个领域中也各有利弊，所以需要依据具体的应用场景和目标来进行合理的选择，并综合运用这些方法，以此实现较优的数据处理和分析效果。

## 6. 聚焦数学建模与大数据分析典型应用案例

### 6.1. 气象预报建模

在大数据时代，气候变化领域涌现各式各样的数据集。这些数据集有着多种模式，在表示形式、分布状况、比例大小、密度高低等方面也存在着差异。故，对大数据挖掘与大数据分析方法的需求越来越迫切，需要依赖其进行数据挖掘，提供可靠的决策，对后期的研究以及相关性分析给予重要帮助。故，当前对大数据挖掘与分析的需求日益迫切，它能为数据挖掘提供支撑、为决策提供可靠依据，并辅助后续研究及相关分析。

例如 Sentinel-2 马赛克大气底部图像，该图像具有 18,881 个单独的  $100 \times 100$  公里图块。该研究的结果支持空间明确的海草和海洋生态系统核算，并能进一步协助政策制定、蓝碳信用和所有必要的金融投资。该研究使用随机森林机器学习算法来评估和分析所需的数据。Béjar 等人使用具有 rHEALPix 的地球观测数据立方体实施了离散的全球网格系统(DGGS)，以实现不同空间数据的有效集成[11]。Nikolaev 等人，利用来自 CanESM2 的 800 年时间序列的一般环流模型模拟数据，使用深度学习方法对其进行训练并使用历史数据进行测试[12]。

### 6.2. 癌症放疗中的数学建模

长期以来，数学建模于癌症放疗领域始终扮演着基石性的角色，凭借诸多应用对癌症治疗的处方拟定、计划编排以及实际实施发挥着不可或缺的指导作用。伴随医疗大数据时代的来临，肿瘤及患者层面的分子数据、成像数据与临床数据呈现出整合之势，这一趋势有望推动癌症治疗朝着更为精准、个性化的方向迈进。在此背景下，数学建模的重要性愈发凸显。它既能深入挖掘整合数据中的关键信息，也能据此构建贴合个体特征的治疗模型，从而为精准化、个性化癌症治疗提供扎实的理论依据与实践指引，已成为推动该领域高质量发展的重要支撑。

在肿瘤治疗领域，数学建模对于肿瘤控制概率(TCP)以及正常组织并发症概率(NTCP)模型的构建与应用有着举足轻重的地位。这两类模型旨在肿瘤根除最大化与周边健康组织伤害最小化之间寻求最优平衡，而达成这一平衡对于优化整个肿瘤治疗策略而言是至关重要的环节。

具体而言，很多学者围绕此开展了诸多相关研究工作。例如，Watanabe 等人通过构建一个相对简洁的数学模型，对肿瘤生长以及其在单次高剂量分数照射下的反应情况进行模拟。在该模型中，运用了如来自线性二次模型(LQ 模型)的辐射敏感性  $\alpha$  指标这类关键生物学参数，同时引入了新的血管生长迟缓因

子  $\theta$ , 以此来预测在实验及临床环境中, 经过治疗后肿瘤体积所发生的变化[13]。Matsuya 等人则采用马尔可夫链蒙特卡洛模拟的方法, 针对 LQ 模型与另外两个微剂量动力学模型(MKM)展开对比分析。经研究发现, 在照射过程中, 损伤修复这一环节起着极为关键的作用, 并且在短剂量输送时间的情况下, 那些未考虑修复因子的模型很可能会高估分次放疗过程中的细胞杀伤效果[14]。

随着该领域的进步, 数学建模的重要性将日益凸显, 从而推动该领域进一步完善, 并将多尺度和多模型医疗数据整合到精确和个性化的癌症治疗方法中。

## 7. 总结

交通大数据分析是数学建模从理论走向城市交通工程实际应用的桥梁, 在整个过程中有着至关重要的作用。

在数据预处理环节, 面对海量且繁杂的交通数据, 需要进行诸如数据清洗, 去除重复、错误以及缺失值过多的数据等操作, 让后续拥堵预测建模的数据基础更加扎实可靠。同时, 通过标准化、归一化等操作, 使不同量级的交通数据能处于同一尺度, 便于模型的处理。特征工程则是挖掘交通数据中有价值的特征, 通过特征提取、选择、构建等, 提炼出那些真正对模型结果有显著影响的关键交通因素, 让模型能够去芜存菁。而模型验证阶段, 借助交叉验证、留出法等手段, 衡量模型的性能优劣, 及时发现过拟合、欠拟合等问题并加以调整, 进而提升模型的准确性、鲁棒性和可解释性。

本文以城市交通拥堵预测为具体应用场景, 完整展示了数学建模的专业过程, 包括假设条件提出、核心方程建立、求解算法设计及模型稳定性证明, 验证了数学建模与交通大数据分析融合的实际价值。未来, 随着大数据技术蓬勃发展, 交通数据能够实现实时采集与处理, 使交通领域数学建模工作可以基于最新数据展开。因果推断等方法的应用有助于突破交通数据表层关联、揭示拥堵形成的深层因果关系, 而轻量化模型则增强了在资源受限的中小城市交通管理场景下的适用性。

在复杂交通系统建模中, 比如城市交通网络的动态模拟、交通流演变预测等, 交通大数据与建模的融合能精准刻画系统机制; 在交通应急调度、城市路网规划等决策支持场景中, 二者结合也能助力决策者依据精准预测快速做出科学决策, 推动交通领域数学建模不断向数据驱动精准预测迈进。

## 参考文献

- [1] Bargagna, F., De Santi, L.A., Martini, N., Genovesi, D., Favilli, B., Vergaro, G., *et al.* (2023) Bayesian Convolutional Neural Networks in Medical Imaging Classification: A Promising Solution for Deep Learning Limits in Data Scarcity Scenarios. *Journal of Digital Imaging*, **36**, 2567-2577. <https://doi.org/10.1007/s10278-023-00897-8>
- [2] Wang, J., Yin, Y. and Wei, L. (2025) Modeling Public Opinion Dynamics in Social Networks Using a GAN-SEIR Framework. *Social Network Analysis and Mining*, **15**, Article No. 40. <https://doi.org/10.1007/s13278-025-01426-x>
- [3] Jamarani, A., Haddadi, S., Sarvizadeh, R., Haghi Kashani, M., Akbari, M. and Moradi, S. (2024) Big Data and Predictive Analytics: A Systematic Review of Applications. *Artificial Intelligence Review*, **57**, Article No. 176. <https://doi.org/10.1007/s10462-024-10811-5>
- [4] Ohue, M., Yasuo, N. and Takata, M. (2024) Innovations in Mathematical Modeling, AI, and Optimization Techniques. *The Journal of Supercomputing*, **81**, Article No. 340. <https://doi.org/10.1007/s11227-024-06861-9>
- [5] Weng, W. (2024) Artificial Intelligence—Mathematical Modeling. In: Weng, W., Ed., *A Beginner's Guide to Informatics and Artificial Intelligence*, Springer, 29-38. [https://doi.org/10.1007/978-981-97-1477-3\\_4](https://doi.org/10.1007/978-981-97-1477-3_4)
- [6] Tosi, D., Kokaj, R. and Rocchetti, M. (2024) 15 Years of Big Data: A Systematic Literature Review. *Journal of Big Data*, **11**, Article No. 73. <https://doi.org/10.1186/s40537-024-00914-9>
- [7] Pyar, K. (2024). Predictive Analytics System Using Big Data Framework. 2024 *IEEE Conference on Computer Applications (ICCA)*, Yangon, 16-16 March 2024, 1-6. <https://doi.org/10.1109/icca62361.2024.10532805>
- [8] Chaudhary, Y. and Pathak, H. (2025) Role of Machine Learning for Big Data Applications. In: Nedjah, N., *et al.*, Eds., *Proceedings of the International Conference on Smart Systems and Advanced Computing*, Springer, 223-235. [https://doi.org/10.1007/978-3-031-40905-9\\_23](https://doi.org/10.1007/978-3-031-40905-9_23)

- 
- [9] Lawrance, J.U., Jesudhasan, J.V.N. and Thampiraj Rittammal, J.B. (2024) Parallel Fuzzy C-Means Clustering Based Big Data Anonymization Using Hadoop Mapreduce. *Wireless Personal Communications*, **135**, 2103-2130. <https://doi.org/10.1007/s11277-024-11101-7>
- [10] Kanimozhi, A. and Vimala, N. (2023) Adaptive Weighted Support Vector Machine Classification Method for Privacy Preserving in Cloud over Big Data Using Hadoop Framework. *Multimedia Tools and Applications*, **83**, 3879-3893. <https://doi.org/10.1007/s11042-023-15825-9>
- [11] Béjar, R., Lacasta, J., Lopez-Pellicer, F.J. and Noguera-Iso, J. (2023) Discrete Global Grid Systems with Quadrangular Cells as Reference Frameworks for the Current Generation of Earth Observation Data Cubes. *Environmental Modelling & Software*, **162**, Article ID: 105656. <https://doi.org/10.1016/j.envsoft.2023.105656>
- [12] Nikolaev, A., Richter, I. and Sadowski, P. (2020) Deep Learning for Climate Models of the Atlantic Ocean. *AAAI Spring Symposium: MLPS*, Stanford, 23-25 March 2020. <http://chfps.cima.fcen.uba.ar/>
- [13] Watanabe, Y., Dahlman, E.L., Leder, K.Z. and Hui, S.K. (2016) A Mathematical Model of Tumor Growth and Its Response to Single Irradiation. *Theoretical Biology and Medical Modelling*, **13**, Article No. 6. <https://doi.org/10.1186/s12976-016-0032-7>
- [14] Matsuya, Y., Kimura, T. and Date, H. (2017) Markov Chain Monte Carlo Analysis for the Selection of a Cell-Killing Model under High-Dose-Rate Irradiation. *Medical Physics*, **44**, 5522-5532. <https://doi.org/10.1002/mp.12508>