分类号：
学　号：20222508001

密　　级：公开
单位代码：10759

石河子大学

硕士学位论文

基于数据分解和深度学习的 $PM_{2.5}$ 预测研究与系统实现

学位申请人　曾韬
指导教师　徐丽萍 教授

申请学位类别　专业硕士
专业名称　电子信息
研究领域　计算机技术
所在学院　信息科学与技术学院

中国·新疆·石河子
2025 年 5 月

分类号：　　密　　级：公开
学　号：20222508001　　单位代码：10759

石河子大学

硕 士 学 位 论 文

基于数据分解和深度学习的 $PM_{2.5}$ 预测研究与系统实现

学 位 申 请 人	曾韬
指 导 教 师	徐丽萍 教授
申 请 学 位 类 别	专业硕士
专 业 名 称	电子信息
研 究 领 域	计算机技术
所 在 学 院	信息科学与技术学院

中国·新疆·石河子

2025 年 5 月

Research and System Implementation of $PM_{2.5}$ Prediction Based on Data Decomposition and Deep Learning

A Dissertation Submitted to

Shihezi University

In Partial Fulfillment of the Requirements

for the Degree of

Master of Engineering

By

Zeng Tao

Electronic Information

Dissertation Supervisor: Prof.Xu Li-Ping

May, 2025

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所呈交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：曾韦韵　　时间： 2025 年 5 月 26 日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：曾韦韵　　时间： 2025 年 5 月 26 日

导师签名：蒋丽萍　　时间： 2025 年 5 月 26 日

摘要

随着工业化和城市化进程的加快，$PM_{2.5}$作为空气污染的重要组成部分，对人类健康和环境安全构成了严重威胁。精准预测$PM_{2.5}$浓度变化对于环境管理、公共健康保护以及政策制定等具有重要意义。然而，$PM_{2.5}$数据的固有非线性特性对现有预测方法的精度构成挑战。本文以提高$PM_{2.5}$预测准确性为目标，从数据分解、分解与集成预测、时序相关性建模等方面进行$PM_{2.5}$浓度预测方法研究。主要研究内容包括：

（1）基于一次分解的$PM_{2.5}$预测方法研究。鉴于$PM_{2.5}$随时间变化且受多重因素影响，表现出显著的非线性特征，本文采用CEEMDAN方法对$PM_{2.5}$数据进行分解，旨在降低数据复杂性并提取多尺度特征，从而提升预测精度。在此基础上验证了CEEMDAN在削弱$PM_{2.5}$序列非线性及提取子序列特征方面的有效性。进一步，提出将相近PE值的子序列进行K-means聚类并融合，从而构建了CKP-LSTM模型。与LSTM模型相比，R^2提升了0.014，MAE和RMSE降低了1.16和2.84。同时，实验结果表明高频序列是影响模型精度的关键因素。因此，提出了对高频序列实施二次分解的策略，以期进一步优化预测效能。

（2）基于二次分解的$PM_{2.5}$预测方法研究。针对高频序列的复杂性，本文提出采用自适应分解的WVMD方法进行二次分解，并在此基础上构建出本文所提出的预测模型CKP-WVMD-LSTM。二次分解技术实验表明，WVMD在高频序列的二次分解中展现出显著优势。通过对比实验、消融实验的分析，本文所提出的模型在预测性能上实现了显著提升，与LSTM模型相比，该模型在R^2上提升了0.027，MAE和RMSE分别降低了2.96和5.72。与他人工作的对比中，验证了所提模型的可靠性和有效性。此外，时间和空间泛化性的实验，证明了该模型具有在短期中精准预测及长期中捕捉其变化趋势的能力，并展现出适用于不同地区$PM_{2.5}$浓度的预测任务的潜力。

（3）$PM_{2.5}$浓度预测系统的设计与实现。采用前后端分离的架构模式实现$PM_{2.5}$浓度预测系统。前端使用 Vue 框架结合Echarts可视化库进行开发，后端采用Django框架，并通过MySQL数据库进行数据存储和管理。系统提供了便捷的数据上传和模型训练功能，并可通过模型预测$PM_{2.5}$，然后以图表的方式展示预测结果，使用户能直观地了解$PM_{2.5}$浓度变化趋势。

关键词：$PM_{2.5}$预测；数据分解；分解与集成；长短期记忆神经网络；鲸鱼优化算法

Abstract

With the acceleration of industrialization and urbanization, $PM_{2.5}$, as an important component of air pollution, poses a serious threat to human health and environmental safety. Accurate prediction of $PM_{2.5}$ concentration changes is important for environmental management, public health protection, and policy formulation. However, the inherent nonlinear nature of $PM_{2.5}$ data poses a challenge to the accuracy of existing prediction methods. In this thesis, with the goal of improving $PM_{2.5}$ prediction accuracy, we conduct research on $PM_{2.5}$ concentration prediction methods in terms of data decomposition, decomposition and integration prediction, and time-series correlation modeling. The main research content includes:

（1） Research on $PM_{2.5}$ prediction method based on primary decomposition. Given that $PM_{2.5}$ varies over time and is affected by multiple factors, exhibiting significant nonlinear characteristics, this thesis adopts the CEEMDAN method to decompose $PM_{2.5}$ data, aiming at reducing data complexity and extracting multiscale features so as to improve prediction accuracy. On this basis, the effectiveness of CEEMDAN in weakening the nonlinearity of $PM_{2.5}$ series and extracting subsequence features is verified. Further, K-means clustering and fusion of subsequences with similar PE values are proposed to construct the CKP-LSTM model. Compared with the LSTM model, the R^2 is improved by 0.014, and the MAE and RMSE are reduced by 1.16 and 2.84. Meanwhile, the experimental results show that the high-frequency sequences are the key factors affecting the accuracy of the model. Therefore, the strategy of implementing quadratic decomposition for high-frequency sequences is proposed in order to further optimize the prediction efficacy.

（2） Research on $PM_{2.5}$ prediction method based on quadratic decomposition. In view of the complexity of high-frequency sequences, this thesis proposes to use the WVMD method of adaptive decomposition for secondary decomposition, and based on this, the prediction model CKP-WVMD-LSTM proposed in this thesis is constructed. Experiments of secondary decomposition technique show that WVMD demonstrates a significant advantage in secondary decomposition of high-frequency sequences. Through the analysis of comparative experiments and ablation experiments, the proposed model in this thesis achieves a significant improvement in prediction performance, which improves 0.027 in R^2, and reduces 2.96 and 5.72 in MAE and RMSE, respectively, compared with the LSTM model. The reliability and validity of the proposed model are verified in the comparisons with the work of others. In addition, experiments on temporal and spatial generalizability demonstrate the model's ability to accurately predict in the short term and capture its trends in the long term, and show the potential to be applicable to the task of predicting $PM_{2.5}$ concentrations in different regions.

（3） Design and implementation of $PM_{2.5}$ concentration prediction system. The $PM_{2.5}$ concentration prediction system is realized by adopting the architecture pattern of front-end and back-end separation. The

front-end is developed using the Vue framework combined with the Echarts visualization library, and the back-end adopts the Django framework with data storage and management through a MySQL database. The system provides convenient data uploading and model training functions, and can predict $PM_{2.5}$ through the model, and then display the prediction results in the form of charts, so that users can visualize the trend of $PM_{2.5}$ concentration changes.

Key words: $PM_{2.5}$ prediction; Data decomposition; Decomposition and integration; Long short-term memory neural network; Whale Optimization Algorithm

目录

第 1 章 绪论

1.1 研究背景与意义

在全球工业化和城市化迅速推进的背景下，空气污染已成为一个备受关注的热点问题。习近平总书记在二十大报告中强调，要大力推进环境污染防治，加强污染物协同控制，基本消除重污染天气。$PM_{2.5}$作为空气污染的主要组成部分[1]，对人类健康和环境安全造成了严重影响[2, 3]。$PM_{2.5}$作为细颗粒物的一种[4, 5]，因其极小的体积（直径小于或等于 2.5 微米），能深入肺部甚至血液，对人体健康构成直接威胁，包括增加呼吸系统疾病、心血管疾病、肺癌以及早产和低出生体重儿等健康风险。据世界卫生组织（WHO）统计，$PM_{2.5}$每年造成上百万人的过早死亡。此外，$PM_{2.5}$还通过散射和吸收太阳辐射，影响地球的能量平衡，进而加剧气候变化，减少能见度[6]，影响交通安全，并对自然生态系统造成深远影响[7]，如减少农作物产量、破坏森林生态等。

面对这一严峻挑战，全球范围内对空气质量管理和$PM_{2.5}$治理的重视程度不断提升。我国作为世界上最大的发展中国家，近年来在环境保护领域取得了显著成就，特别是在$PM_{2.5}$治理方面。通过修订《环境空气标准》[8]、实施严格的排放控制、推广清洁能源、加强工业污染治理、建设大规模的空气质量监测[9]网络等措施，有效降低了全国范围内的$PM_{2.5}$浓度，显著改善了空气质量。然而，由于地域辽阔、发展不平衡、污染源复杂多样等因素，部分地区$PM_{2.5}$污染问题依然严峻，特别是在冬季采暖期、重工业集中区和城市群区域，$PM_{2.5}$污染仍时有发生，对当地居民的健康和生活质量构成严重威胁。

$PM_{2.5}$的来源具有多样性及复杂性，将其彻底消除几乎是不可能的，最佳的应对措施是对其精准的预测[10]。$PM_{2.5}$预测不仅是科学研究的前沿课题，也是实现环境精准治理、保障公众健康、促进社会经济绿色转型的迫切需求。精准的$PM_{2.5}$预测可以为当地政府和公众提供精确的空气质量监测和预警服务[11]，助力政府科学规划交通出行、精准制定$PM_{2.5}$污染防控工作的环保政策，并高效实施应急响应措施，以确保民众健康、提升生活环境质量及维护生态环境的持续稳定, 实现环境效益和经济效益的双赢。

近年来，鉴于数据量的急剧增长与人工智能技术的飞速发展，为$PM_{2.5}$预测提供了新的技术手段和解决方案。众多领域纷纷采纳人工智能技术，其在大气污染研究领域的应用价值也日益凸显。在数据支撑上，$PM_{2.5}$实时监测数据为学者对$PM_{2.5}$预测研究奠定了数据基础。从时间尺度上来看，$PM_{2.5}$预测可分为短期、中期和长期三类。目前的预测模型中，短期预测通常能实现较高的精度，但预测时长有限且稳定性不高。中期和长

期预测虽时间长，但精度往往欠佳，不能实际应用。因此，构建一个既能实现短期精准预测，又能在长期预测中捕捉 $PM_{2.5}$ 变化趋势的稳定模型，具有很高的应用价值。

1.2 国内外研究现状

自二十世纪五十年代起，英国等国家便率先展开了空气质量的研究[12]，而我国在此领域的研究则稍晚起步。我国的大气污染物浓度预测工作始于二十世纪七十年代首次全国环保会议之后，而到了九十年代，$PM_{2.5}$ 预测逐渐成为了研究热点，并在后续年份中持续发展并趋于完善。目前，$PM_{2.5}$ 预测领域已涌现出多种预测模型，包括数值模拟模型、统计学模型以及人工智能模型等。这些技术不仅显著提升了 $PM_{2.5}$ 预测的效率和准确性，同时也促进了相关预测模型的不断进步。接下来，将深入探讨这些技术在 $PM_{2.5}$ 预测中的具体应用与研究现状。

1.2.1 基于数值模拟模型的研究现状

数值模拟模型以大气物理和化学原理为基础，综合考虑气象条件、扩散传输过程和污染源排放，具有强大的物理依据，能够提供详尽的大气污染模拟和解释[13, 14]，可通过模拟空气污染物的物理传输和化学反应来预测 $PM_{2.5}$ 浓度。王茜等人[15]采用空气质量模型（CMAQ）对上海市十个国控站点检测数据进行预报，并采用学习型线性回归方法对结果进行修正，提高了预测的准确性。Zhang 等人[16]基于气象数据和 ERA Interim 再分析数据，使用在线耦合的 WRF-Chem 模型和 GFS 数据，评估了每小时和每天 $PM_{2.5}$ 浓度的可预测性，为实时空气质量预报提供了参考。付高平等人[17]用 CMAQ，以污染物数据为基础，融合气象资料、污染源和地形地貌等数据，模拟成都市区各季度的 $PM_{2.5}$ 浓度，这种多源数据融合的方法为模型预测提供了更为全面和准确的信息支持。Li 等人[18]利用混合单粒子拉格朗日积分轨迹（HYSPLIT）模型，结合卫星数据和 $PM_{2.5}$ 监测数据，定量分析了不同排放条件下各城市燃烧生物质上风向 $PM_{2.5}$ 的区域输送，揭示了 $PM_{2.5}$ 的区域传输特征，为制定跨区域污染防控策略提供了科学依据。

然而，数值模拟模型在预测结果中产生的差异仍然是一个不容忽视的问题。这些差异往往源于输入数据（如排放源和排放量）的不准确性、模型参数的设置、气象条件的复杂性以及化学反应机制的复杂性等多种因素[4, 5]。这些限制因素使得模型在实时预测方面的表现难以达到理想状态。

1.2.2 基于统计学模型的研究现状

统计学模型大多依赖于数学方程或模拟技术来描述空气污染的演变，包括线性回归

模型、移动平均自回归模型（ARIMA）和灰色模型等[19]。Zhang 等人[20]融入 NO_2和 EVI 的地理加权回归（GWR）模型更有效地估计全国范围内的 $PM_{2.5}$，可以解释相应 $PM_{2.5}$ 质量浓度变化的约 87%。Chen 等人[21]基于对长沙市 $PM_{2.5}$ 历史数据和相关天气信息的分析，建立了一个多元线性回归模型（MLR）来预测 $PM_{2.5}$ 浓度。Ma 等人[22]开发了一个国家尺度 GWR 模型，以融合卫星 AOD 为主要预测因子，估算中国每日 $PM_{2.5}$ 浓度。高丽霞等人[23]运用灰色模型（GM（1，1））， 结合 2015-2021 年宁夏数据，预测空气污染物趋势。结果显示 SO2、PM10 等六项指标均趋平稳下降，对宁夏大气环境治理有指导意义。Gao 等人[24]采用平滑指数法对中国台湾地区未来六小时的污染物浓度进行预测。Wang 等人[25]使 ARIMA 预测了从加州空气资源委员会提取的 $PM_{2.5}$ 数据，结果表明，季节性 ARIMA 模型可以成为预测空气污染的有效方法。

尽管统计学模型广泛用于污染物浓度预测，但多基于单变量线性时间序列数据，而 $PM_{2.5}$ 等污染物数据复杂多变，具非线性、突变性，统计学模型在处理此类非线性关系上力不从心。为了更好的学习这种非线性关系，人工智能模型受到广泛关注。

1.2.3 基于深度学习模型的研究现状

根据模型结构和学习方法不同，常用的人工智能模型可以分为传统机器学习模型和深度学习模型。例如，支持向量机（SVR）、随机森林（RF）等传统机器学习模型可以进行 $PM_{2.5}$ 预测。Hou 等人[26]在 SVR 模型中使用高斯核函数以及 k 倍交叉验证和网格搜索方法来获得最优参数，对 PM_{10} 和 $PM_{2.5}$ 的预测具有良好的泛化能力。杨立娟等人[27]构建的包含 AOD 和其他辅助变量的 2 层随机森林模型可有效获取近地面 $PM_{2.5}$ 浓度的空间分布。Ma 等人[28]利用极端梯度提升算法（XGBoost）对上海市 $PM_{2.5}$ 浓度进行了预测，实验结果表明，该模型在预测 $PM_{2.5}$ 浓度方面展现出了较高的精确度。Niu 等人[29]利用最小二乘支持向量机对广州和兰州两地 $PM_{2.5}$ 浓度数据进行预测，结果显示最小二乘支持向量机预测效果比 BP 人工神经网络预测效果好。然而，这些模型不能自动提取数据特征且学习长期依赖关系的能力有限。深度学习[30]的特征工程机制及强大的序列学习能力可有效缓解传统机器学习模型所存在的问题。

在深度学习模型中，循环神经网络（RNN）[31]是一种具有循环连接结构的神经网络，因为其特有结构非常符合相关大气环境的动态特性，经常用来模拟空气污染物分布的时间演化。Feng 等人[31]结合 RF 和 RNN 来分析和预测杭州空气污染物的下一个 24 小时 $PM_{2.5}$ 浓度。然而，传统 RNN 在处理长序列时可能会面临梯度消失和梯度爆炸等问题，为此，长短期记忆神经网络（LSTM）和门控循环神经网络（GRU）等 RNN 的改进型被广泛使用来解决这些问题。作为 RNN 的代表网络，引入细胞状态和门控机制的 LSTM 是 $PM_{2.5}$ 预测任务中最常用的深度学习方法。Ren 等人[32]以前一时刻多变量时间序列，使用 LSTM 预测当前时刻 $PM_{2.5}$ 浓度，实验结果表明，LSTM 的预测性能优于 SVR、

ARIMA 等传统回归模型。由于 LSTM 只能处理过去时刻的信息，Du 等人[33]基于双向长短期记忆神经网络（BiLSTM）和一维卷积神经网络（1D-CNN），利用时间序列过去和未来的信息预测 $PM_{2.5}$ 浓度，预测精度优于 LSTM 和 ARIMA 等基准模型。Li 等人[34]使用了结构更为简单和学习效率更高的 GRU 来预测南京市 $PM_{2.5}$ 浓度。

$PM_{2.5}$ 浓度时间序列因其显著的非平稳性、非线性等特征而极具复杂性[35, 36]，这导致单一模型在处理时难以全面把握其内在规律，进而阻碍了预测精度的提升。一系列研究表明，混合模型的预测精度常常优于单一模型。Li 等人[6]提出的 1D-CNN 和 LSTM 混合模型，可以同时学习到 $PM_{2.5}$ 浓度的时空相关性和长期依赖性，进一步提高预测精度。Zhang 等人[37]提出的基于残差神经网络（ResNet）和卷积长短期记忆神经网络（ConvLSTM）的混合模型，能够在提取多个城市污染物和气象数据空间特征的同时还能提取到相应的高维数据时空特征，从而在一段时间内准确预测目标城市未来的 $PM_{2.5}$ 浓度。赵鹏飞等人[38]构建了一个融合注意力机制的门控循环神经网络（Attention+GRU）模型，用于预测北京市 2010 年至 2014 年的 $PM_{2.5}$ 浓度数据，结果显示，融合了注意力机制的模型在预测精度上显著优于未融合的模型。Ding 等人[39]基于时空相关性的混合 CNN-LSTM 模型来预测北京 $PM_{2.5}$ 日浓度，实验表明考虑时空相关性的模型的预测精度优于没有时空相关性的相同模型。Muthukumar 等人[40]利用各种大气污染物的遥感卫星图像图形结合卷积网络（GCN）和 ConvLSTM 来学习 $PM_{2.5}$ 在空间和时间相关性上的模式，误差结果显示，与该领域现有的研究相比，有了显著的改进。Thiruchelvam L 等人[41]将注意力机制融合到 LSTM 搭建混合模型，实验发现该模型比 ARIMA 模型预测精度高。

1.2.4 基于数据分解和深度学习混合模型的研究现状

目前，越来越多的学者采用数据分解算法，通过数据分解算法，能在降低原始序列复杂度的情况下还能获得多尺度的特征信息。通过深入挖掘 $PM_{2.5}$ 序列内部特征，并将其与人工智能模型结合，以实现更精准的预测。相较于单一人工智能模型，这种组合方法显著提升了预测精度。在数据分解方法中，一个典型的方法是将 $PM_{2.5}$ 数据分解为多个子序列，将每个子序列输入到对应的预测模型中，最后将结果进行聚合以获得最终的预测结果，这种方法称为分解与集成方法[42-44]。常用的数据分解算法包括周期趋势分解（STL）、小波变换（WT）和模态分解（MD）等。

对于 STL 的研究，Yuan 等人[45]提出了一种通过基于 STL 的改进的物态启发式算法（DSMS）训练的树突神经元模型（DS-DNM），其中利用 STL 算法进行数据预处理，从原始数据中获得周期分量、趋势分量和残差分量，通过 DSMS 训练的 DNM 预测残差值，将三组特征量相加以获得预测值，实验发现 DS-DNM 在 $PM_{2.5}$ 浓度预测问题上具有更强的竞争力。WT 是一种卓越的变换分析方法，可以将原始数据进行分解，有利于 $PM_{2.5}$

预测任务。Cheng 等人[46]在 WT 的基础上对原始数据进行分解，然后结合简单的预测器如 ANN 和 SVM 对序列数据进行预测，与传统的 ANN 和 SVM 相比，基于 WT 的预测模型各种评价指标在不同城市的预测中都有明显的提高。Qiao 等人[20]提出了一种基于 WT、LSTM 和堆叠式自动编码器（SAE）的模型，利用 WT 将 $PM_{2.5}$ 时间序列分解为低频和高频序列，再基于 SAE-LSTM 对分解后的序列进行预测，最后重构所有低频序列和高频序列预测结果，预测性能优于其他基线模型。与 WT 相比，模态分解（MD）方法原将始 $PM_{2.5}$ 复杂时间序列数据分解为多个子信号序列，可以能够更好地适应数据的非线性、非平稳特性，提供更丰富的多尺度信息，逐渐受到研究人员的广泛关注。Teng 等人[47]采用基于经验模态分解（EMD）的混合预测模型能够准确预测短期（6 h 以内）$PM_{2.5}$ 浓度，并捕捉到长期（6-24 h）变化趋势。对于 EMD 衍生出的一些变体，使用集成经验模态分解（EEMD）和广义回归神经网络（GRNN）来预测 $PM_{2.5}$ 序列[48]，精度高于单个 GRNN 模型。Niu 等人[49]对 EMD 改进，提出了基于互补式组合经验模态分解（CEEMD）的 SVR 模型，实验表明该模型有更高的预测精度。自适应噪声完备集合经验模态分解（CEEMDAN）在每个阶段对噪声自适应叠加，解决了 EEMD 中重构误差高的问题。Zhang 等人[42]利用 CEEMDAN 对 $PM_{2.5}$ 进行分解，然后采用模糊聚类（C-means）对本征模态函数（IMF）进行分组，最后应用 LSTM 进行 $PM_{2.5}$ 预测，效果比单一 LSTM 显著提高。相较之下，变分模态分解（VMD）可以有效克服模态混叠且具有更严格的数学理论基础。Zhang 等人[42]提出的 VMD-BiLSTM 混合模型可以显著提高 $PM_{2.5}$ 预测精度，但仅仅依靠人工选择 VMD 的参数，效率并不高。针对此问题，Shi 等人[50]采用灰狼优化算法（GWO）来优化 VMD 的参数，不仅提高了参数选择的效率，还提升了预测的精度。Chu 等人[51]提出了一种基于 VMD、状态转换模拟退火算法和 SVR 的城市 $PM_{2.5}$ 短期混合预测模型，实验发现，与现有的预测方法相比，该模型具有较强的鲁棒性。

一次分解似乎能够达到不错的预测效果，但分解后可能仍然存在难以预测的高复杂度序列[44]，很多学者认为需要结合预测模型进行二次分解的研究，以期提取序列内部更多的特征信息。Wu 等人[52]提出一种结合小波分解（MD）和 VMD 的二次分解方法来预测每日空气质量指数（AQI），结果表明，经过 VMD 二次分解的聚合模型效果好于仅 MD 分解的聚合模型。Dong 等人[3]提出一种基于 CEEMDAN 和 VMD 的二次分解技术与 LSTM 结合预测每日的 $PM_{2.5}$ 浓度，结果显示该模型的稳定性和精度都有显著提升。Wu 等人[53]将 CEEMDAN 分解后最大 PE 的子序列用灰狼优化器（GWO）优化过的 VMD 进行二次分解，结合基于注意力机制的双向长短期记忆网络用于 $PM_{2.5}$ 的小时预测。与其他模型相比，该模型可以更准确地预测每小时 $PM_{2.5}$ 浓度。

在梳理和总结国内外关于 $PM_{2.5}$ 预测研究的现状后，发现众多学者已从多元化的角度和研究方向对 $PM_{2.5}$ 浓度的预测进行了深入的分析与探讨，并提出了多种预测方法。

然而，当前的预测研究工作仍面临着一系列的问题与挑战。针对此，将需要进一步研究的问题归纳如下：

（1）随着人工智能技术得到快速发展，目前 $PM_{2.5}$ 预测模型大多以深度学习模型为主，且对于 $PM_{2.5}$ 浓度这样的时间序列数据，深度学习中的时间序列模型显得更有优势。

（2）$PM_{2.5}$ 预测任务受其数据本身非线性、不稳定性的影响，导致预测精度不高。因此，将数据分解策略与深度学习技术相结合，削弱其非线性的同时提供多尺度的特征，能够提高 $PM_{2.5}$ 预测的准确性。

（3）混合的模型效果往往好于单个模型。可以基于分解与集成思想，对分解后的数据分别用深度学习时间序列模型捕捉数据的非线性和时序特征，构建一个混合模型以提高预测精度及稳定性。

1.3 研究内容

本文通过分析 $PM_{2.5}$ 数据特性找出 $PM_{2.5}$ 预测的难点，以 $PM_{2.5}$ 数据为基础，同时考虑气象条件对 $PM_{2.5}$ 浓度的影响，把多变量时间序列数据的复杂非线性特性作为切入点展开研究，基于分解与集成思想，构建基于数据分解策略和深度学习技术相结合的 $PM_{2.5}$ 预测模型并将其应用到预测系统中。具体研究内容如下：

（1）基于一次分解的$PM_{2.5}$预测方法研究。$PM_{2.5}$浓度具有复杂的变化趋势，这对模型的预测精度提出了挑战。针对$PM_{2.5}$数据的非线性特征，通过CEEMDAN分解有效的降低$PM_{2.5}$浓度数据的非平稳性及非线性程度，并提供多尺度特征信息，有利于提高的模型的预测精度。

（2）基于二次分解的 $PM_{2.5}$ 预测方法研究。通过对 CEEMDAN 分解后的序列分析，发现高频序列会影响预测性能，针对于高频序列提出 WVMD 自适应分解方法，并在此基础上构建本文提出的模型，通过对高频序列的分解，进一步提升模型性能。

（3）$PM_{2.5}$ 预测系统的设计与实现。采用 B/S 架构，基于 Django、Vue、MySQL 等技术开发 $PM_{2.5}$ 预测系统，将所提出的模型应用到系统中，实现数据管理、模型管理、$PM_{2.5}$ 预测等功能。

1.4 技术路线

本文基于$PM_{2.5}$数据和气象数据，融合数据分解技术和深度学习方法构建$PM_{2.5}$预测模型，并将所构建的模型应用到$PM_{2.5}$预测系统中。本文的总体技术路线图如图1-1所示。