

实现中国碳达峰目标的挑战与前景分析的研究

摘要

随着气候变化成为全球紧迫议题，中国在实现碳中和目标的过程中仍面临着诸多挑战，本文通过建立数学模型，分析预测碳排放趋势，评估各因素对实现碳中和目标的影响，为政府制定减排政策提供科学依据，提高中国碳达峰目标的现实性和可行性，助力全球应对气候变化。

对于问题一，需要分析全球碳排放量、自然灾害发生量、地球地表温度变化三者间的关系。本文首先对全球碳排放量、自然灾害发生量、地球地表温度三个附录文件，进行整合、分析与数据清洗，然后绘制了三者数据随时间序列变化的趋势图，然后通过 **Pearson 相关性分析** 的方法得出碳排放量、自然灾害发生量与地球地表温度之间的相关性系数，对三者关系进行简单分析，并运用 **三次样条插值函数**，通过 **最小二乘法拟合** 得到三者的插值拟合图得到两两之间的关系。最后，在上述模型经过假设检验后，总结出全球碳排放量、自然灾害发生量与地球地表温度之间的关系为：**全球碳排放量与地球地表温度变化呈正相关，与自然灾害发生量呈负相关**。对于地球地表温度变化与自然灾害发生量，两者间并不能得到明显且单一的线性关系。

对于问题二，需要将一氧化氮排放量、甲烷排放量、碳排放量三者进行比较，然后建立合适的评价模型，评价用碳排放量作为应对全球气候变化，减缓全球温室气体排放速度的指标的合理性。本文在问题一所建立模型的基础上，先对数据进行预处理，然后截取 2010 年到 2020 年间的有效数据进行数据可视化，运用 **Pearson 相关系数** 得到四项排放量间的 **Pearson 相关系数热力图**，经过假设检验，最终确认碳排放量与全球温室气体排放量的 **Pearson 相关系数** 的绝对值最大，说明在一氧化氮排放量、甲烷排放量、碳排放量三者中，**碳排放量与全球温室气体排放量的相关性最高**，故而用碳排放量作为应对全球气候变化，减缓全球温室气体排放速度的指标是具有一定的合理性的。

对于问题三，需要分析影响国内碳排放变化的主要影响因素和贡献率，并建立合适的碳排放量模型并预测 2030 年碳达峰目标实现的可行性。首先对数据进行预处理，得到影响国内碳排放变化的 8 个影响因素，基于前文所建立的 **Pearson 相关系数模型** 进行特征选择，去除国内二氧化硫排放量后，选出影响国内碳排放变化的 7 个主要影响因素为：农村碳排放量、人口密度、人均可支配收入、本专科毕业生数、能源消费总量、公共图书馆业机构数、工业企业单位数，将这 7 个变量作为预测变量，将国内碳排放量作为响应变量，基于随机森林模型进行多元回归分析，构建回归模型预测国内碳排放量。对于 7 个预测变量，再基于 **LSTM 模型** 分别构建其各自的时间序列预测模型，预测各自 2018 到 2032 年的数据，最后将数据代入多元回归分析所得的回归模型，得到自 2018 到 2032 年国内碳排放量的预测值，根据预测结果可知我国在 2030 年后碳排放量达到峰值，即我国在 2030 年实现碳达峰的目标是可行的。

对于问题四,人工智能技术在科技减碳方面具有重要价值。基于前三问的解答,以及当前研究领域的发展,我们提出了以下几点建议:可以利用 AI 赋能,优化能源结构、提高能效、建立智能交通系统、实现建筑智能化、发展精准农业技术、与碳捕捉和存储技术结合,以及智能数据分析和监测等。这些技术手段能够推动清洁能源的开发利用、提高能效、优化交通流量、实现建筑节能、推动环保农业实践、优化碳捕获和储存流程以及持续跟踪和报告碳排放情况。综合运用这些技术,将有助于推动全球绿色低碳发展,为应对气候变化提供重要支持。未来,人工智能在科技减碳方面的作用将更加显著。

关键字: 0-1 标准化 Pearson 相关系数 t 检验 三次样条插值 随机森林回归 多元回归分析

一、问题重述

1.1 问题背景

气候变化的影响深远且广泛，对粮食生产构成威胁，并增加了洪灾的风险。为应对这一全球性挑战，各国纷纷签署国际公约，共同致力于实现碳中和目标。作为温室气体排放的主要国家之一，中国面临着重大的压力和挑战。然而，我们必须清醒地认识到，中国在实现碳中和目标的过程中，仍然面临着能源结构、产业结构等方面的严峻挑战。实现 2030 年碳达峰目标需要付出长期而艰苦的努力。本文基于现有数据，运用科学的数学模型，对中国未来的碳排放量趋势进行了预测，并对各个因素对中国碳中和目标的影响进行了评估。这将为我们制定更加科学、合理的政策和措施提供有力的支撑，为实现碳中和目标奠定坚实的基础。

1.2 问题要求

问题 1 分析全球碳排放量、自然灾害发生量与地球地表温度之间的关系。

问题 2 建立评价模型，证明碳排放量作为应对全球气候变化，减缓全球温室气体排放速度的指标的合理性。

问题 3 分析碳排放量的主要影响因素，预测 2030 年碳达峰目标实现的可行性。

问题 4 结合人工智能等技术与本文所建数学模型的评价、预测结果，提出合理科学的科技减碳方案。

二、问题分析

2.1 问题一分析

对于问题一，要求分析本文首先对全球碳排放量、自然灾害发生量、地球地表温度三者间的关系。本文首先对三者的附录数据数据清洗、集成、归约与变换，通过 0 - 1 标准化对数据进行归一化处理，截取三者从 2000 年到 2020 年的相关数据，据此绘制了全球碳排放量、自然灾害发生量与地球地表温度的面板数据随时间序列变化的趋势图、增量数据随时间序列变化的趋势图，然后通过 Pearson 相关性分析的方法得出全球碳排放量、自然灾害发生量与地球地表温度之间的相关性系数，对全球碳排放量、自然灾害发生量、地球地表温度三者间关系进行简单分析，并运用三次样条插值函数，通过最小二乘法拟合得到三者的插值拟合图，分析得出两两之间的关系。最后，在上述模型经过假设检验后，总结得出全球碳排放量、自然灾害发生量与地球地表温度之间的关系为：全球碳排放量与自然灾害发生量、地球地表温度变化均具有较强的线性相关性，其中全

球碳排放量与地球地表温度变化呈正相关，与自然灾害发生量呈负相关。对于地球地表温度变化与自然灾害发生量，两者间并不能得到明显且单一的线性关系。

2.2 问题二分析

对于问题二，要求将一氧化氮排放量、甲烷排放量、碳排放量三者进行比较，然后建立合适的评价模型，评价用碳排放量作为应对全球气候变化，减缓全球温室气体排放速度的指标的合理性。本文在问题一所建立模型的基础上，先对附录中的一氧化氮排放量、甲烷排放量、碳排放量和全球温室气体排放量文件进行数据预处理，对四个文件中的有效数据进行数据标准化，然后在一氧化氮排放量、甲烷排放量、碳排放量三者排放量三者数据中，选取 1990 年到 2020 年间 30 年的数据进行数据可视化后，进行分析比较。然后截取 2010 年到 2020 年间的有效数据，运用 Pearson 相关系数得到四项排放量间的 Pearson 相关系数热力图，对结果进行假设检验，若通过假设检验，且碳排放量与全球温室气体排放量的 Pearson 相关系数的绝对值最大，则判断用碳排放量作为应对全球气候变化，减缓全球温室气体排放速度的指标是具有一定的合理性的，反之则不太合理。

2.3 问题三分析

对于问题三，要求分析影响国内碳排放变化的主要影响因素和贡献率，并建立合适的碳排放量模型并预测 2030 年碳达峰目标实现的可行性。首先对数据进行预处理，得到影响国内碳排放变化的影响因素，通过前文所建立的 Pearson 相关系数模型，计算出各种影响因素与国内碳排放量的相关系数，通过比较相关性大小来进行特征选择，得到影响国内碳排放变化的主要影响因素。而后基于随机森林思想的组合分类器设计，采用 Bagging 算法进行集成学习，然后通过多元回归分析，将国内作为响应变量 F ，选出的主要影响因素作为预测变量 Q ，构建回归模型。对于选出的主要影响因素，分别构建其各自的时间序列预测模型，预测各自 2018 到 2035 年的数据，最后将数据代入多元回归分析所得的回归模型，得到自 2018 到 2035 年国内碳排放量的预测值，根据预测结果判断我国在 2030 年实现碳达峰的目标是否可行。

2.4 问题四分析

对于问题四，在探讨人工智能技术在科技减碳方面的建议时，我们需要综合考虑前文所提及的多个方面，并结合当前人工智能技术的最新进展和减碳领域的实际需求，提出具有针对性和可操作性的建议。

三、模型假设

为简化问题，本文做出以下假设：

- 全国的碳排放量视作表格中各区的碳排放量的求和。
- 假设历史国内碳排放量数据能够代表未来的变化趋势。
- 数据均准确无误。

四、符号说明

符号	说明	单位
k	原始数据	
k^*	离差标准化后的数据	
C_p	Pearson 相关系数	
N	变量的样本数目	
X_i	碳排放量	klt
Y_i	自然灾害发生量	
Z_i	地球地表温度数据	$^{\circ}C$

五、问题一的模型的建立和求解

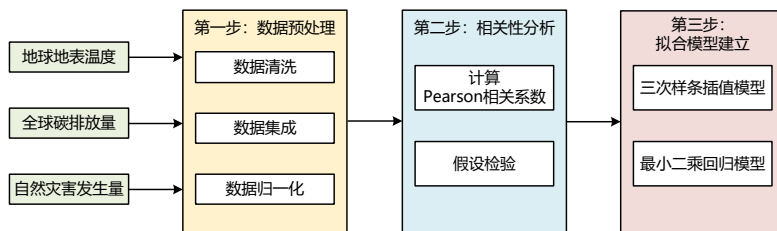


图 1 问题一求解流程图

5.1 模型建立

5.1.1 数据预处理

对于问题一，本文首先要求分析全球碳排放量、自然灾害发生量、地球地表温度三者间的关系。为此，本文对三者的附录数据进行了清洗、集成、归约与变换，通过 0-1 标准化对数据进行归一化处理，并截取了从 2000 年到 2020 年的相关数据。据此，我们

绘制了全球碳排放量、自然灾害发生量与地球地表温度的面板数据随时间序列变化的趋势图。随后，通过 **Pearson** 相关性分析的方法，我们得出了全球碳排放量、自然灾害发生量与地球地表温度之间的相关性系数，并对这三者间的关系进行了简单分析。此外，我们还运用了三次样条插值函数，并通过最小二乘法拟合得到了三者的插值拟合图，进一步分析了它们两两之间的关系。最后，在经过假设检验后，我们总结得出：全球碳排放量与自然灾害发生量、地球地表温度变化均具有较强的线性相关性，其中全球碳排放量与地球地表温度变化呈正相关，与自然灾害发生量呈负相关。然而，地球地表温度变化与自然灾害发生量之间并未呈现出明显且单一的线性关系。

$$k^* = \frac{k - \min k}{\max k - \min k} \tag{1}$$

式(1)中， k^* 表示离差标准化后的数据， $\min k$ 为样本数据的最小值， $\max k$ 为样本数据的最大值

5.2 初步分析与判断

对处理后的数据，利用 **Matlab** 绘制出全球碳排放量、自然灾害发生量与地球地表温度的面板数据随时间变化趋势图和全球碳排放变化量、自然灾害发生变化量、地球地表温度变化量，三者的差分数据随时间序列变化的趋势图，如下图所示：

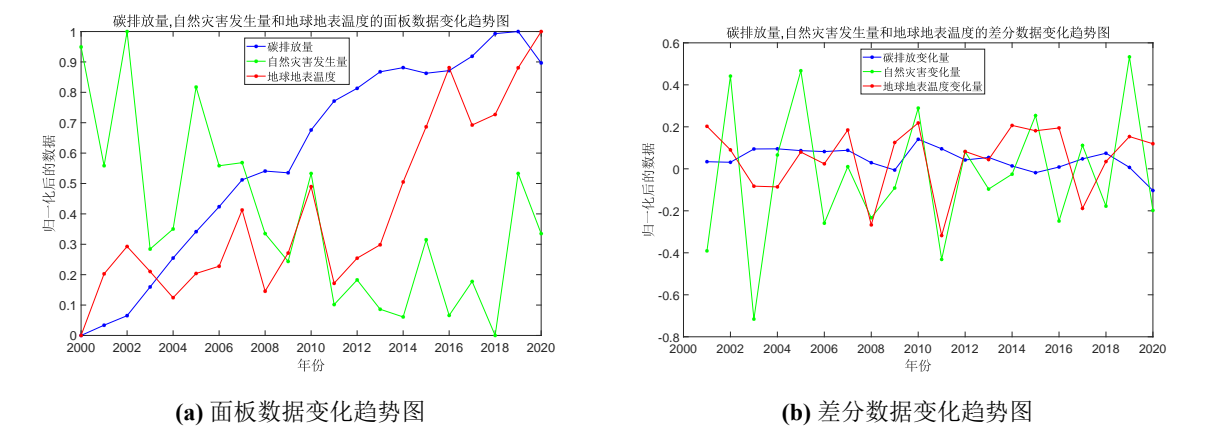


图 2

分析图2a可知从 2000 年到 2020 年，全球碳排放量与地球地表温度整体呈上升趋势，初步推断两者为正相关，而全球碳排放量与地球地表温度整体呈下降趋势，因此初步推断自然灾害发生量与全球碳排放量与、球地表温度分别为负相关。因而可以初步推断全球碳排放量、自然灾害发生量与地球地表温度这三个变量，两两之间具有线性关系。

分析图2b可知从 2000 年到 2020 年，自然灾害发生量的变化量与地球地表温度变化量，两个差分序列具有较高的一致性，即地球地表温度的变化量下降（上升）时，自然灾害发生量的变化量也下降（上升）。初步推断地球地表温度的变化速度与自然灾害发生量的变化速度可能具有一定的关联性。

上述由变量随时间变化的趋势图，可以横向对比，初步推断全球碳排放量、自然灾害发生量与地球地表温度三者间的关系，但结果可能偏离度较大，无法准确表述三者间的关系，因此本文采用 Pearson 相关性分析，进一步分析三者间的关系。

5.3 Pearson 相关系数模型建立

5.3.1 Pearson 相关系数

Pearson 相关系数可以用来度量两个变量 a 和 b 之间的线性相关关系，其值介于-1 到 1 之间。这里定义两个变量 a 和 b 之间的 Pearson 相关系数 $C_p(a, b)$:

$$C_p(a, b) = \frac{cov(a, b)}{\sigma_a \sigma_b} = \frac{\sum_i^N (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i^N (a_i - \bar{a})^2} \sqrt{\sum_i^N (b_i - \bar{b})^2}} \quad (2)$$

式(2)中， N 表示变量的样本数目即 2000 年到 2020 年的年份数， a_i 和 b_i 分别表示 a 和 b 的第 i 个取值， \bar{a} 和 \bar{b} 分别是 a 和 b 的均值。

$cov(a, b)$ 表示变量 a 和变量 b 间的协方差，用来度量两个变量间变化趋势是否一致，协方差公式如下：

$$cov(a, b) = \frac{1}{N} \sum_i^N (a_i - \bar{a})(b_i - \bar{b}) \quad (3)$$

σ_a 和 σ_b 是 a 和 b 的标准差，用来衡量数据集中数据点的离散程度或波动性的统计量，以变量 a 为例，标准差公式如下：

$$\sigma_a = \sqrt{\frac{\sum_i^N (a_i - \bar{a})^2}{N}} \quad (4)$$

5.3.2 假设检验

对于 Pearson 相关性系数 $C_p(a, b)$ ，其绝对值越接近于 1，证明对应两个变量的相关性越好，以此逻辑，本文构建 Pearson 相关性分析模型，但对于 Pearson 相关系数的求解结果，需要通过假设检验来验证 Pearson 相关性系数是否显著不同于 0。本文运用 t 检验进行假设检验，步骤如下：

Step1: 设置零假设 H_0

相关性系数为 0，即两个变量之间不存在线性相关关系。

Step2: 设置备择假设 H_1

相关性系数不为 0，即两个变量之间存在显著的线性相关关系。

Step3: 计算 t 统计量

$$t = C_p(a, b) \sqrt{\frac{N - 2}{1 - [C_p(a, b)]^2}} \quad (5)$$

式(5)中, 表示相关性系数, N 表示变量的样本数目。**Step4:** 设定置信水平为 95%, 则显著性水平 $\alpha = 1 - 95\% = 0.05$

根据样本量 n 和显著性水平 α 确定 t 分布的自由度和临界值, 依据 t 分布的自由度和临界值选择拒绝域的边界。

Step5: 检验 t 统计量 t 值是否落在 t 分布的拒绝域内

如果 t 统计量落在拒绝域内, 则拒绝零假设, 认为两个变量之间存在显著的线性相关关系; 否则接受零假设, 认为两个变量之间不存在显著的线性相关关系。一般来说, 如果 p 值小于显著性水平 α , 就可以拒绝原假设。 p 值是在给定原假设下观察到的统计量或者更极端统计量的概率。

5.4 Pearson 相关系数模型求解

5.4.1 判断全球碳排放量、自然灾害发生量、地球地表温度间的线性相关关系

Step1: 计算 Pearson 相关系数依照所建立的 Pearson 相关性分析模型, 对全球碳排放量、自然灾害发生量、地球地表温度三个变量进行不重复的两两组合 C_3^2 , 由式(2)分别计算它们的 Pearson 相关系数为:

$$C_p(Carbon, Disaster) = -0.711478242$$

$$C_p(Carbon, Temperature) = 0.735122127$$

$$C_p(Disaster, Temperature) = -0.355346491$$

将所得 Pearson 相关系数数据绘制成热力图3, 图中, 颜色越深代表两个变量间的相关性越强, 方格中数据为两个变量间的相关系数 $C_p(a, b)$ 。

Step2: t 检验

本文显著性水平 α 取 0.05, 在此对三者间 Pearson 相关系数进行 t 检验, 由式(2)计算得出三者间各自的 t 统计量后, 查找 t 分布表得到各组的 p 值, 并与显著性水平 α 比较, 得到表达式如下:

$$p(Carbon, Disaster) = 0.000298552 < \alpha$$

$$p(Carbon, Temperature) = 0.000146843 < \alpha$$

$$p(Disaster, Temperature) = 0.113926282 > \alpha$$

则假设检验结果为: 拒绝全球碳排放量与自然灾害发生量不存在显著的线性相关关系这一零假设, 拒绝全球碳排放量与地球地表温度不存在显著的线性相关关系这一零假设, 接受地球地表温度与自然灾害发生量不存在显著的线性相关关系这一零假设。

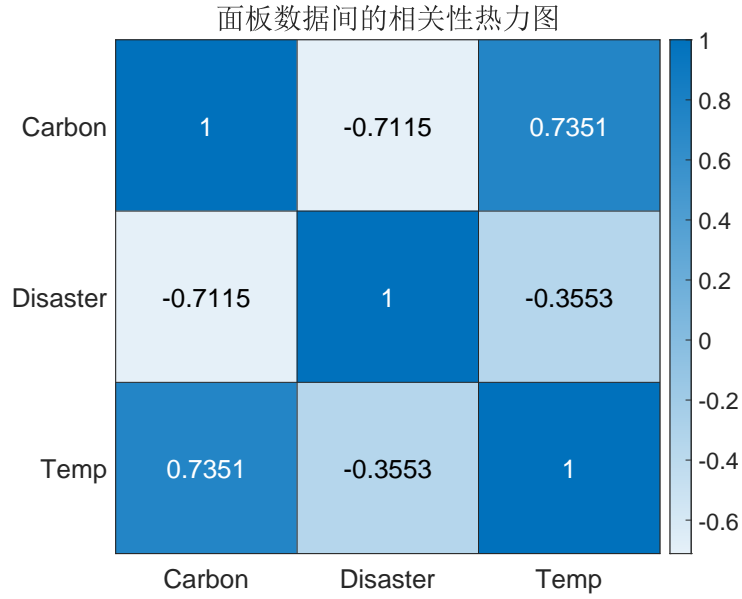


图 3 全球碳排放量、自然灾害发生量与地球地表温度 **Pearson** 相关系数热力图

Step3: 总结

依据图3及假设检验，可知全球碳排放量与地球地表温度具有较强的正相关性，若全球碳排放量上升，地球地表温度也会上升，反之亦然。全球碳排放量与自然灾害发生量具有较强的负相关性，若全球碳排放量降低，自然灾害发生量反而会上升。而自然灾害发生量与地球地表温度不存在显著的线性相关关系。故初步分析所得推断部分正确，更进一步推断为：全球碳排放量与地球地表温度具有正相关性，与自然灾害发生量具有负相关性，自然灾害发生量与地球地表温度不具有相关性。

5.4.2 判断全球碳排放变化量、自然灾害发生变化量、地球地表温度变化量之间的线性相关关系

Step1: 计算 Pearson 相关系数

具体步骤同 5.3.1: 依照所建立的 Pearson 相关性分析模型，对全球碳排放变化量、自然灾害发生变化量、地球地表温度变化量三个变量进行不重复的两两组合 C_3^2 ，由式(2)分别计算它们的 Pearson 相关系数为：

$$C_p(\text{Carbon}, \text{Disaster}) = -0.03748184$$

$$C_p(\text{Carbon}, \text{Temperature}) = -0.27056164$$

$$C_p(\text{Disaster}, \text{Temperature}) = 0.361841976$$

其中, *Carbon* 表示全球碳排放量, *Disaster* 表示自然灾害发生量, *Temperature* 表示地球地表温度。将所得 Pearson 相关系数数据绘制成热力图4, 图中, 颜色越深代表两个变量间的相关性越强, 方格中数据为两个变量间的相关系数, 数据可视化如下:

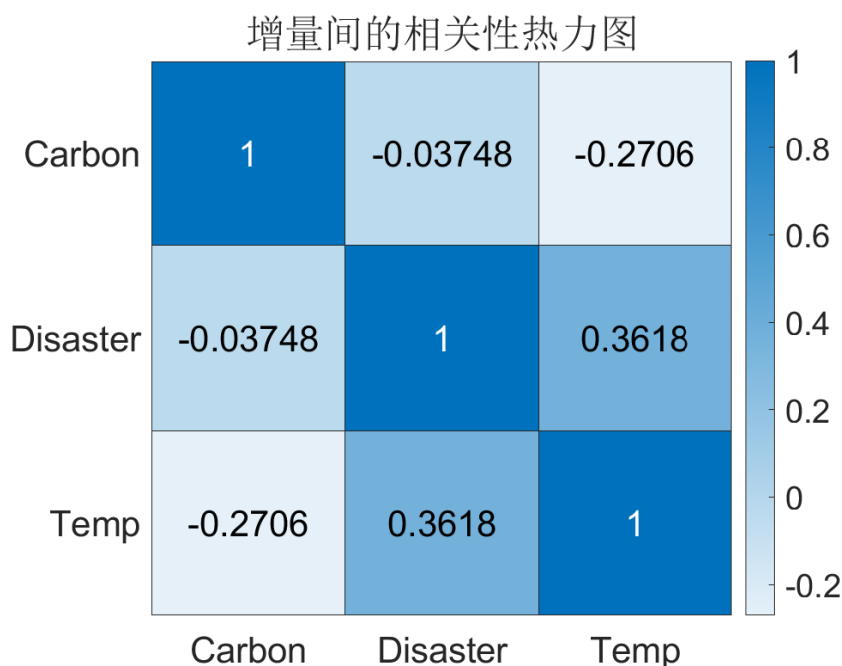


图 4 全球碳排放变化量、自然灾害发生变化量与地球地表温度变化量 Pearson 相关系数热力图

Step2: t 检验

本文显著性水平 α 取 0.05, 在此对三者间 Pearson 相关系数进行 t 检验, 由式(5)计算得出三者间各自的 t 统计量后, 查找 t 分布表得到各组的 p 值, 并与显著性水平 α 比较, 得到表达式如下:

$$p(\text{Carbon}, \text{Disaster}) = 0.875335395 > \alpha$$

$$p(\text{Carbon}, \text{Temperature}) = 0.248599595 > \alpha$$

$$p(\text{Disaster}, \text{Temperature}) = 0.116956267 > \alpha$$

则假设检验结果为: 接受全球碳排放变化量与自然灾害发生变化量不存在显著的线性相关关系这一零假设, 接受全球碳排放变化量与地球地表温度变化量不存在显著的线性相关关系这一零假设, 接受地球地表温度变化量与自然灾害发生变化量不存在显著的线性相关关系这一零假设。

Step3: 总结

由假设检验可知全球碳排放变化量、自然灾害发生变化量、地球地表温度变化量两两之间无显著的线性相关关系，初步分析所得推断错误，后续过程不再考虑三者间的关系。

5.5 插值拟合模型建立与求解

5.5.1 三次样条插值模型建立

常见的插值方法包括线性插值、拉格朗日插值、牛顿插值、样条插值等，其中三次样条插值，相较于其他常见插值方法，通过三次多项式片段来拟合数据，从而获得连续且具有较高光滑度的插值曲线。故而本文采用三次样条插值建立数学模型，来判断全球碳排放量、自然灾害发生量与地球地表温度三者间的关系。

三次样条插值是将插值区间内的数据点与相邻区间的数据点之间连接的三次多项式称为样条函数，从而使整个插值区间内的插值曲线是由多个三次多项式片段组成的。本文用表示变量的样本数目，在问题一中即表示 2000 年到 2020 年的年份数，设 X_i 为碳排放量数据， Y_i 为自然灾害发生量数据， Z_i 为地球地表温度数据。设数据点为 P_i ，问题一中需进行三次样条插值的数据点类型为三维数据点：(X_i, Y_i, Z_i)。

三次样条插值拟合模型建立步骤如下：

Step1: 将 N 个数据点 $P_0, P_1, P_2, \dots, P_{N-1}$ 的插值区间 $[x_0, x_{N-1}]$ 分成 $N-1$ 个子区间 $[x_0, x_1], [x_1, x_2], \dots, [x_{N-2}, x_{N-1}]$ ，每个子区间内使用一个三次多项式进行插值。

Step2: 对于每个子区间 $[x_{i-1}, x_i]$ ，其中， i 表示第 i 个子区间，三次多项式的表达式为：

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, i = 0, 1, \dots, N - 1. \quad (6)$$

则 N 个区间对应的三次函数 $S(x)$ 的数学表达式如下：

$$S(x) = \begin{cases} s_0(x) = a_0 + b_0(x - x_0) + c_0(x - x_0)^2 + d_0(x - x_0)^3, & \text{if } x_0 \leq x \leq x_1 \\ s_1(x) = a_1 + b_1(x - x_1) + c_1(x - x_1)^2 + d_1(x - x_1)^3, & \text{if } x_1 \leq x \leq x_2 \\ s_2(x) = a_2 + b_2(x - x_2) + c_2(x - x_2)^2 + d_2(x - x_2)^3, & \text{if } x_2 \leq x \leq x_3 \\ \vdots \\ s_{N-1}(x) = a_{N-1} + b_{N-1}(x - x_{N-1}) + c_{N-1}(x - x_{N-1})^2 + d_{N-1}(x - x_{N-1})^3, & \text{if } x_{N-1} \leq x \leq x_N \end{cases} \quad (7)$$

Step3: 设置以下条件

$$\left\{ \begin{array}{l} S(x_i) = y_i, i = 0, 1, \dots, N. \\ S_{i-1}(x_i) = S_i(x_i). \\ S'_{i-1}(x_i) = S'_i(x_i). \\ S''_{i-1}(x_i) = S''_i(x_i). \\ S''_0(x_0) = 0, S''_{N-1}(x_N) = 0. \\ S'_0(x_0) = C_1, S'_{N-1}(x_N) = C_2. \end{array} \right. \quad (8)$$

式(8)作为前提条件,保证了 N 段三次函数必须穿过所有已知节点,且所有节点(除第一节点与最后一个节点)处 0、1、2 阶连续,同时使三次多项式具有自由边界与固定边界条件。其中 C_i 表示不同的任意常数

Step4: 得出未知参数方程组

先分别求出函数 $S(x)$ 的一阶导数与二阶导数如下:

$$\left\{ \begin{array}{l} S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3. \\ S'_i(x) = b_i + 2c_i(x - x_i) + 3d_i(x - x_i)^2. \\ S''_i(x) = 2c_i + 6d_i(x - x_i). \end{array} \right.$$

设 $h_i = x_{i+1} - x_i, m_i = S''_i(x_i) = 2c_i$, 由式(3)推导可得:

$$\left\{ \begin{array}{l} a_i = y_i \\ a_i + h_i b_i + h_i^2 c_i + h_i^3 d_i = y_{i+1} \\ b_i + 2h_i c_i + 3h_i^2 d_i = b_{i+1} \\ 2c_i + 6h_i d_i = 2c_{i+1} \\ c_i = \frac{m_i}{2} \\ d_i = \frac{m_{i+1} - m_i}{6h_i} \\ b_i = \frac{y_{i+1} - y_i}{h_i} - \frac{h_i}{2} m_i - \frac{h_i}{6} (m_{i+1} - m_i) \\ h_i m_i + 2(h_i + h_{i+1}) m_{i+1} + h_{i+1} m_{i+2} = 6 \left[\frac{y_{i+2} - y_{i+1}}{h_{i+1}} - \frac{y_{i+1} - y_i}{h_i} \right] \end{array} \right. \quad (9)$$

Step5: 对数据降维

由于问题一中数据点 (X_i, Y_i, Z_i) 为三维数据,故在此对其降维,将三维坐标在空间中的点 (X_i, Y_i, Z_i) 连接起来形成一维向量 $[X_i, Y_i, Z_i]$, 从而方便处理。同时,在插值拟合结束后,需要依照拟合曲面的平滑性对插值曲面的插值效果进行评估。

5.5.2 三次样条插值模型求解

对于 (X_N, Y_N, Z_N) 这一三维数据类型, 本文通过上述所建立的三次样条插值模型, 首先对三维数据进行降维处理, 后代入三次多项式(6):

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, i = 0, 1, \dots, N - 1.$$

通过条件方程组(3)

$$\left\{ \begin{array}{l} S(x_i) = y_i, i = 0, 1, \dots, N. \\ S_{i-1}(x_i) = S_i(x_i). \\ S'_{i-1}(x_i) = S'_i(x_i). \\ S''_{i-1}(x_i) = S''_i(x_i). \\ S''_0(x_0) = 0, S''_{N-1}(x_N) = 0. \\ S'_0(x_0) = C_1, S'_{N-1}(x_N) = C_2. \end{array} \right.$$

绘制插值拟合曲面坐标图的不同方向视图如图5所示：

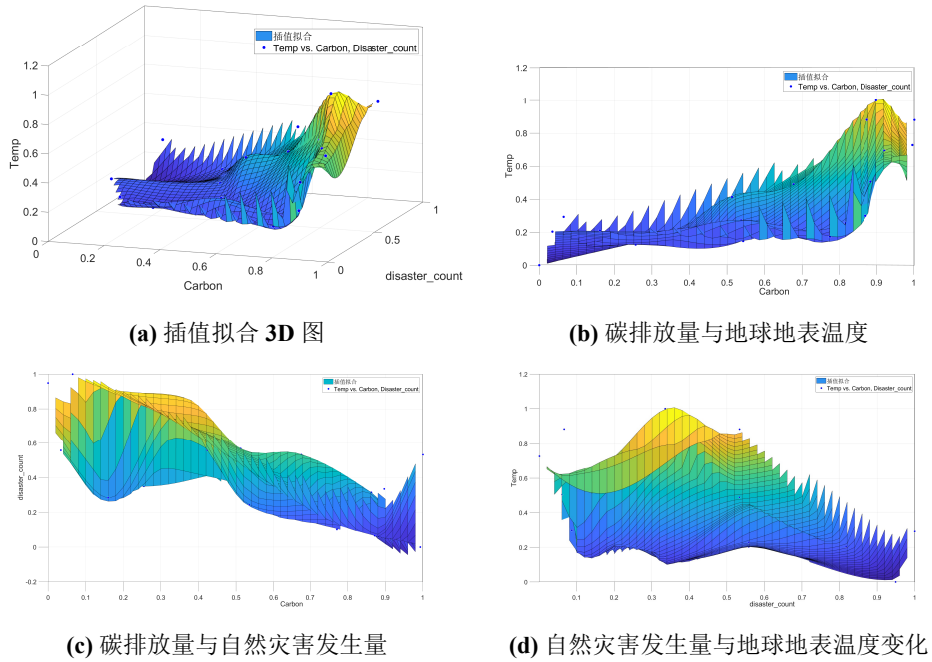


图 5 不同方向视图

由图不难看出，随着碳排放量的升高，地球地表温度变化具有升高的趋势，自然灾害的发生量具有下降的趋势。对于自然灾害发生量和地球地表温度变化，并不能由图示看出明显的线性相关性，故插值拟合的结果与 Pearson 相关系数模型所得结果相互佐证，最终得出全球碳排放量、自然灾害发生量与地球地表温度之间的关系为：全球碳排放量与自然灾害发生量、地球地表温度变化均具有较强的线性相关性，其中全球碳排放量与地球地表温度变化呈正相关，与自然灾害发生量呈负相关。对于地球地表温度变化与自然灾害发生量，两者间并不能得到明显且单一的线性关系。

5.6 最小二乘回归模型建立与求解

一元最小二乘回归的一般形式为：

$$f(x) = ax + b$$

自然灾害发生量与碳排放量的关系：

$$y = -0.61039x + 0.744554$$

地球地表温度与碳排放量的关系：

$$y = 0.63013x + 0.040607$$

两者的最小二乘回归拟合图如图6所示：

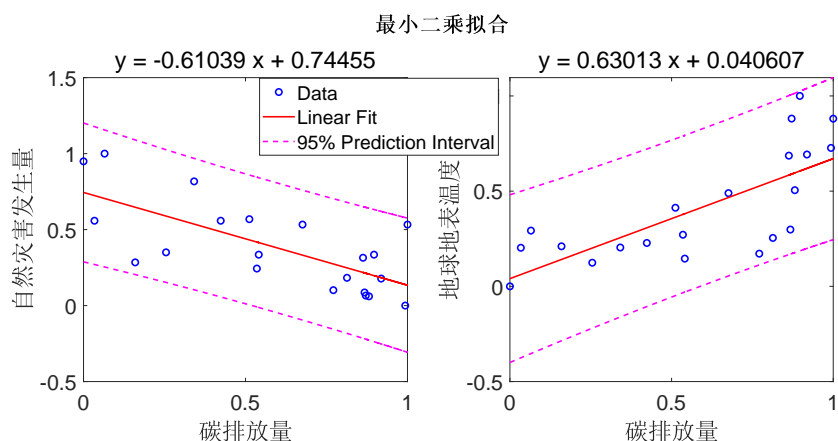


图 6 最小二乘回归拟合图

两者的解释方差都在 95% 以内，可见存在线性关系。

六、问题二的模型的建立和求解

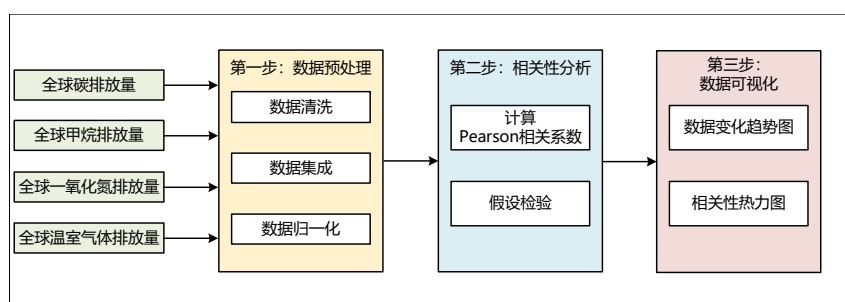


图 7 问题二求解流程图

6.1 数据预处理

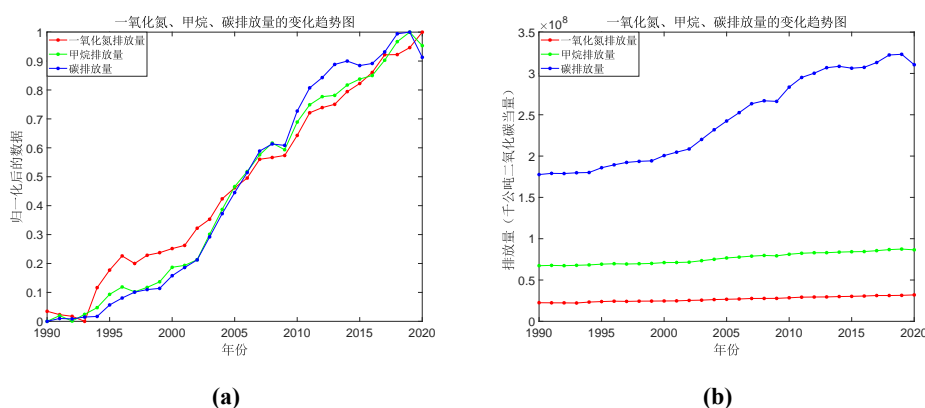
对于全球碳排放量、全球甲烷排放量，全球一氧化氮排放量和全球温室气体排放量四个附录文件，由于其中数据准确性、完整性和一致性都较差，故需要进行数据预处理，以数据清洗、集成、归约、变换等方式提高数据质量。本文通过对全球碳排放量、全球

甲烷排放量，全球一氧化氮排放量和全球温室气体排放量数据进行清洗，去除其中的无效数据，对于有效数据的部分缺失值采用均值插补法进行补全。然后通过数据属性选择的方法，删除与全全球碳排放量、全球甲烷排放量，全球一氧化氮排放量和全球温室气体排放量等有效数据不相关或冗余的属性从而进行数据归约。最后通过数据标准化，将全球碳排放量、全球甲烷排放量，全球一氧化氮排放量和全球温室气体排放量等数据按比例缩放，对数据进行无量纲处理，完成数据变换，运用 0 – 1 标准化方法对原始数据进行线性变换，使结果落在 [0, 1] 区间内，原始数据转换的函数如下：

$$k^* = \frac{k - \min k}{\max k - \min k}$$

6.2 比较全球碳排放量、全球甲烷排放量，全球一氧化氮排放量和全球温室气体排放量

由于问题二要比一氧化氮排放量、甲烷排放量、碳排放量，故本文在进行数据归一化时过程中保留了 1990 年到 2020 年间三者的原始有效数据，并绘制出全球碳排放量、全球甲烷排放量，全球一氧化氮排放量和全球温室气体排放量原始数据随年份变化对比图8a。对于预处理后的数据，本文选择截取其 1990 年到 2020 年间 30 年的数据进行数据可视化，绘制出全球碳排放量、全球甲烷排放量，全球一氧化氮排放量和全球温室气体排放量数据归一化后随年份变化对比图8b，趋势图如下：



分析图8a易得，随着时间推移，全球碳排放量的原始数据远远大于甲烷与一氧化氮的排放量，且三者的排放量都逐年增加。分析图8b可得，在数据归一化后，一氧化氮、甲烷、碳排放量三者各自每一年相较于前一年的增幅大致相同，变相说明了全球温室气体排放量可能是逐年递增的。

6.3 建立评价模型

6.3.1 数据可视化

对于全球碳排放量、全球甲烷排放量，全球一氧化氮排放量和全球温室气体排放量，本文截取四个变量从 2010 年到 2020 年 10 年间的有效数据，如表1所示：

表 1 2010-2020 年的一氧化氮排放量、甲烷排放量、碳排放量与全球温室气体排放量

NO	CH4	C	Global
28483412.62	81197101.5	283467660.8	920453.5038
29250651.27	82400248.01	295140816.4	944926.7947
29426316.02	82965892.06	300299863.1	957645.4161
29538122.92	83058240.37	306940989	969870.6091
29968173.64	83777380.62	308628456.2	978397.9685
30243935.02	84193604.08	306365564.6	973994.7895
30620325.23	84435561.23	307416548.5	974113.2761
31218087.49	85506705.18	313231764.6	989535.9583
31224261.39	86797389.2	322343484.9	1009034.217
31461311.28	87461366.76	323190967	1010968.234
31986264.1	86516874.62	310528346	975822.7197

绘制成如图9所示的变化趋势图：

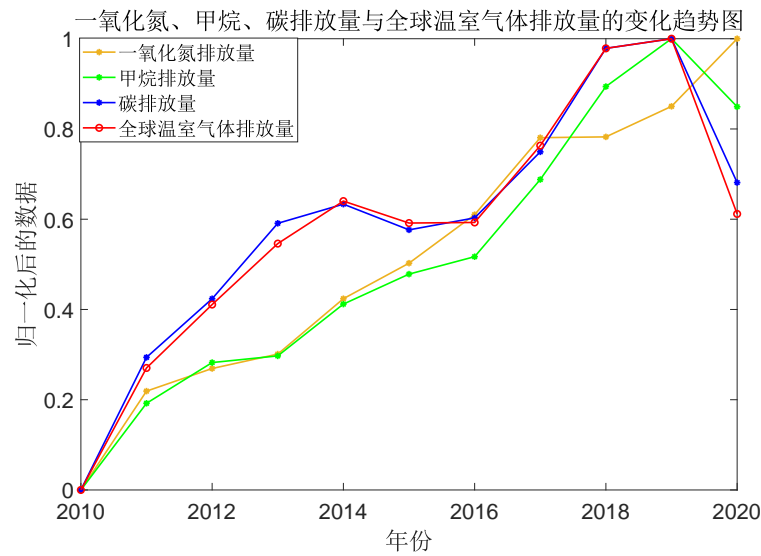


图 9 一氧化氮排放量、甲烷排放量、碳排放量与全球温室气体排放量变化趋势图

分析图9发现，碳排放量与全球温室气体排放量的折线具有高度相似性，相较之下，甲烷排放量、一氧化氮排放量与全球温室气体排放量的折线重合点较少，故初步推断，在图示四个变量中，与全球温室气体排放量相关性最高的点可能是全球碳排放量。现通过 Pearson 相关系数建立评价模型来佐证初步推断。

6.3.2 建立 Pearson 相关系数评价模型

基于问题一所建立的 Pearson 相关系数模型，本文选择用变量的相关性来构建评价模型。通过式(10)计算全球碳排放量、全球甲烷排放量，全球一氧化氮排放量和全球温室气体排放量的 Pearson 相关系数，计算公式如下：

$$C_p(a, b) = \frac{cov(a, b)}{\sigma_a \sigma_b} = \frac{\sum_i^N (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i^N (a_i - \bar{a})^2} \sqrt{\sum_i^N (b_i - \bar{b})^2}} \quad (10)$$

依照所建立的 Pearson 相关性分析模型，对全球碳排放量、全球甲烷排放量，全球一氧化氮排放量和全球温室气体排放量四个变量进行不重复的两两组合 C_3^2 ，由式(10)分别计算它们的 Pearson 相关系数为：

表 2 一氧化氮、甲烷、碳排放量与全球温室气体排放量的 Pearson 相关性矩阵

	NO	CH4	C	Global
NO	1	0.962071	0.851835	0.830449
CH4	0.962071	1	0.925256	0.913004
C	0.851835	0.925256	1	0.995995
Global	0.830449	0.913004	0.995995	1

其中，C 表示全球碳排放量，CH4 表示全球甲烷排放量，NO 表示全球一氧化氮排放量，Global 表示全球温室气体排放量。

将所得 Pearson 相关系数数据绘制成热力图10，图中颜色越深代表两个变量间的相关性越强，方格中数据为两个变量间的相关系数。

对图10简单分析，可以发现，对于全球温室气体排放量，和全球碳排放量、全球甲烷排放量，全球一氧化氮排放量中相关系数最高的为碳排放量，后需进行假设检验证明所得结果的可信度。

6.3.3 t 检验

对计算所得的相关系数进行假设检验，计算各组相关系数的 t 统计量：

$$t = C_p(a, b) \sqrt{\frac{N - 2}{1 - [C_p(a, b)]^2}} \quad (11)$$

检验 t 统计量 t 值是否落在 t 分布的拒绝域内。即检验 p 值是否小于显著性水平 α ，如果 t 统计量落在拒绝域内，则拒绝零假设，认为两个变量之间存在显著的线性相关关系；否则接受零假设，认为两个变量之间不存在显著的线性相关关系。

本文对于所得全球温室气体排放量与全球碳排放量、全球甲烷排放量，全球一氧化氮排放量各自的相关系数进行 t 检验。分别设置零假设为：全球碳排放量与全球温室气

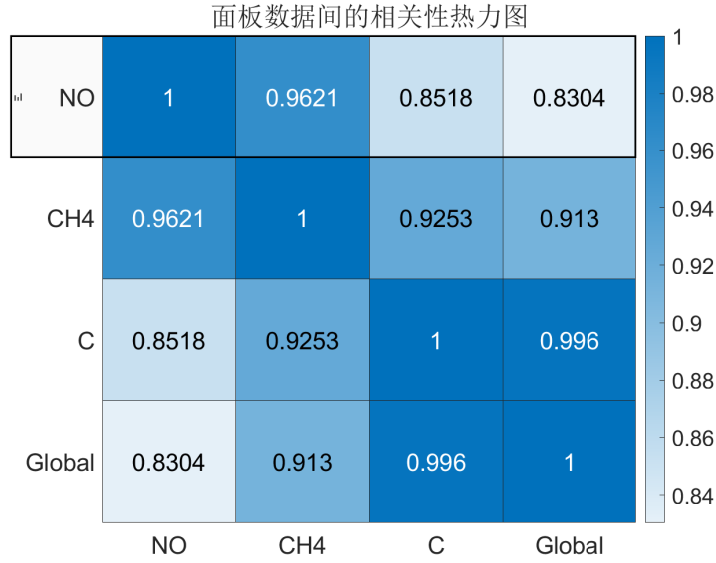


图 10 面板数据间的相关性热力图

体排放量不存在显著的线性相关关系，全球甲烷排放量与全球温室气体排放量的线性相关关系，全球一氧化氮排放量与全球温室气体排放量不存在显著的线性相关关系。

通过式(11)计算得到各组 t 统计量后，查找 t 分布表得到各组的 p 值，并与显著性水平 $\alpha=0.05$ 比较，得到表达式如下：

$$p(C, Global) = 9.48 \times 10^{-11} < \alpha$$

$$p(CH_4, Global) = 8.71 \times 10^{-5} < \alpha$$

$$p(NO, Global) = 0.00155 < \alpha$$

6.3.4 总结

通过假设检验可知，全球温室气体排放量与全球碳排放量、全球甲烷排放量，全球一氧化氮排放量各自的相关系数都通过检验，证明它们各自具有线性相关性。其中全球碳排放量的相关系数最高，故在本文的评价体系下，全球碳排放量对全球温室气体排放量的影响最大，在其中所占权重最高，即用全球碳排放量作为应对全球气候变化，减缓全球温室气体排放速度的指标具有较强的合理性。

七、问题三的模型的建立和求解

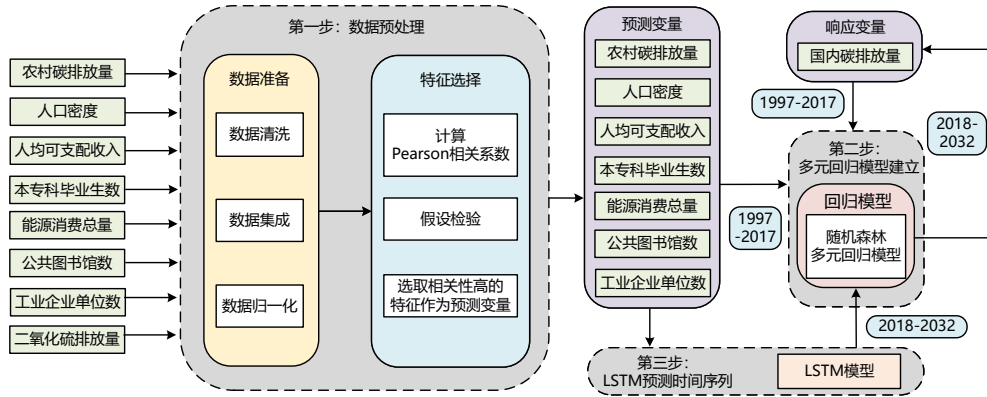


图 11 问题三求解流程图

7.1 数据预处理

观察附录所给数据，选取所有国内相关数据作为对国内碳排放量的影响因素，即本文认为可能影响国内碳排放量的因素有：农村碳排放量、人口密度、人均可支配收入、本专科毕业生数、能源消费总量、公共图书馆业机构数、工业企业单位数。然后对国内碳排放量及选取的影响因素进行数据预处理，基于问题一、二的数据预处理过程，对问题三所需数据进行数据清洗、集成、归约、变换等方式提高数据质量。不同的是，截取问题所需变量数据 1997 到 2017 年共 20 年的有效数据。最后通过数据标准化，将数据按比例缩放，进行无量纲处理，完成数据变换。最后运用 0 – 1 标准化方法，通过式(12)：

$$k^* = \frac{k - \min k}{\max k - \min k} \quad (12)$$

对原始数据进行线性变换，使结果落在 $[0, 1]$ 区间内，对于归一化后的相关数据，绘制国内碳排放量、主观选取的影响因素在 1997 到 2017 年间各自随年份变化的趋势图如图12所示。

分析图12，发现国内二氧化硫排放量折线与国内碳排放量折线差异最大，故初步推断国内二氧化硫排放量可能不是影响国内碳排放量的主要影响因素，而选出的其他影响因素是否是国内碳排放量的主要影响因素难以由趋势图12来判断，需要进一步的模型去进行判断。

7.2 基于 Pearson 相关性分析进行特征选择

基于问题一、二所建模型，本文通过式(13)计算了国内碳排放量与所选取的各影响因素间的 Pearson 相关系数，并经过 t 检验后来判断国内碳排放量的主要影响因素有哪

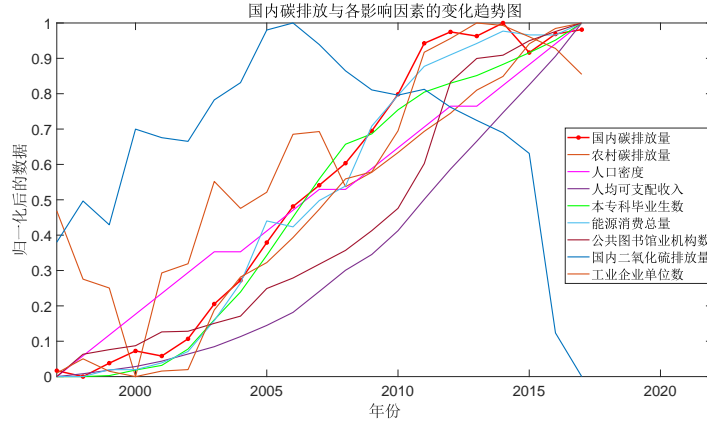


图 12 1997-2017 国内碳排放与各影响因素变化趋势图

些。

$$C_p(a, b) = \frac{cov(a, b)}{\sigma_a \sigma_b} = \frac{\sum_i^N (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i^N (a_i - \bar{a})^2} \sqrt{\sum_i^N (b_i - \bar{b})^2}} \quad (13)$$

计算结果为:

$$C_p(China_Carbon, Countryside_Carbon) = 0.8804$$

$$C_p(China_Carbon, population_density) = 0.9661$$

$$C_p(China_Carbon, income) = 0.9233$$

$$C_p(China_Carbon, graduate) = 0.9915$$

$$C_p(China_Carbon, energy) = 0.9170$$

$$C_p(China_Carbon, library) = 0.9417$$

$$C_p(China_Carbon, SO_2) = -0.0728$$

$$C_p(China_Carbon, enterprise) = 0.9915$$

其中, *China_Carbon* 表示全球碳排放量, *Countryside_Carbon* 表示农村碳排放量, *population_density* 表示人口密度, *income* 表示人均可支配收入, *graduate* 表示本科专科学历毕业生数, *energy* 表示能源消费总量, *library* 表示公共图书馆业机构, *SO₂* 表示国内二氧化硫排放量, *enterprise* 表示工业企业单位数。将所得 Pearson 相关系数绘制成由颜色深浅显示两个变量间相关性的热力图如图13所示:

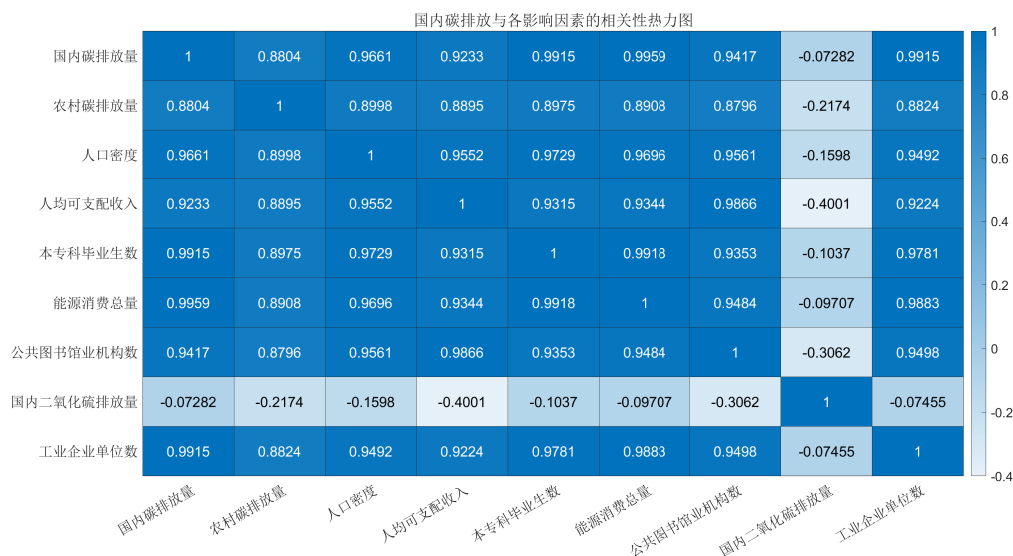


图 13 国内碳排放与各影响因素的相关性热力图

可以提取热力图与国内碳排放量相关的数据，简化为更直观的直方图如图 14所示。

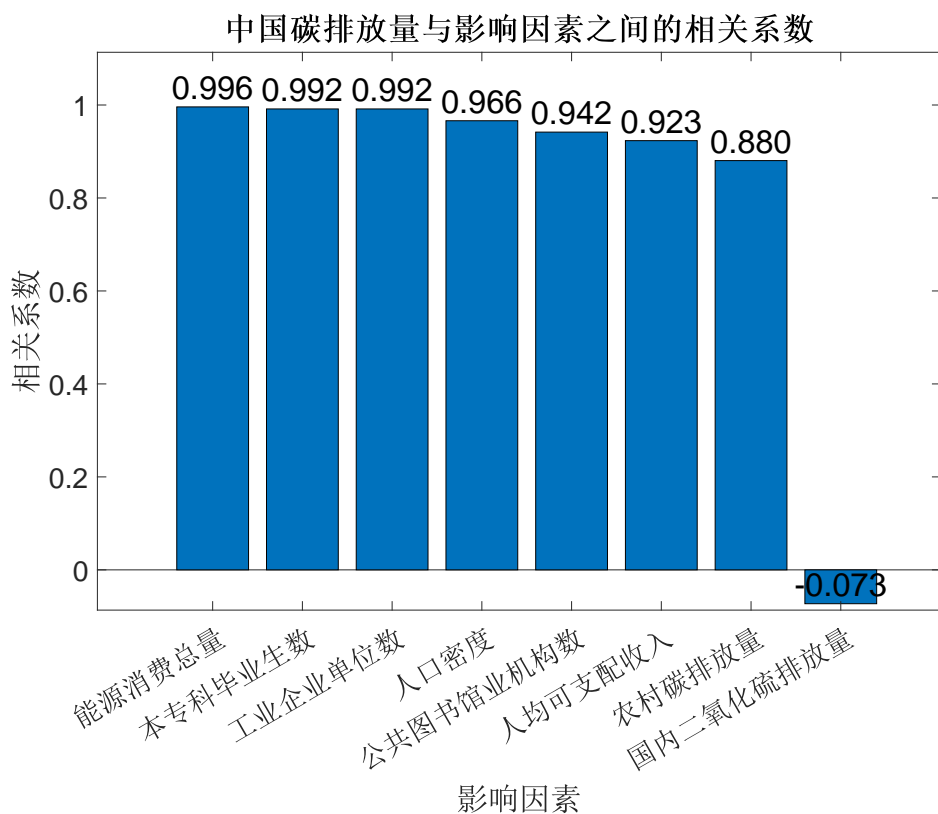


图 14 简化后相关系数直方图

由图不难看出，国内二氧化硫的排放量与国内碳排放量的相关系数的绝对值最低，两者相关性最弱，其余所选取的影响因素与国内碳排放量的相关性较强。但对计算所得的相关系数的合理性，仍需进行假设检验。

设置零假设 H_0 : 相关性系数为 0, 即选取的各影响因素各自与国内碳排放量间都无显著的相关性与备择假设 H_1 : 相关性系数不为 0, 即每两个变量之间存在显著的线性相关关系。由式(14)计算各组相关系数的 t 统计量:

$$t = C_p(a, b) \sqrt{\frac{N - 2}{1 - [C_p(a, b)]^2}} \quad (14)$$

计算后得到:

$$p(\text{China_Carbon}, \text{Countryside_Carbon}) = 0 < \alpha$$

$$p(\text{China_Carbon}, \text{population_density}) = 0 < \alpha$$

$$p(\text{China_Carbon}, \text{income}) = 0 < \alpha$$

$$p(\text{China_Carbon}, \text{graduate}) = 0 < \alpha$$

$$p(\text{China_Carbon}, \text{energy}) = 0 < \alpha$$

$$p(\text{China_Carbon}, \text{library}) = 0 < \alpha$$

$$p(\text{China_Carbon}, \text{SO}_2) = 0.7538 > \alpha$$

$$p(\text{China_Carbon}, \text{enterprise}) = 0 < \alpha$$

故假设检验结果为: 除 SO_2 接受零假设外, 其余所选取的影响因素均拒绝原假设, 即可以判断国内二氧化硫排放量不是国内碳排放量的主要影响因素。

故对于国内碳排放量的主要影响因素进行特征选择, 得主要影响因素有农村碳排放量, 人口密度, 人均可支配收入, 本科专科毕业生数, 能源消费总量, 公共图书馆业机构, 工业企业单位数 7 个特征变量。

7.3 基于随机森林模型的多元回归

7.3.1 随机森林的多元回归模型的建立

随机森林模型的实质是一个包含多个决策树的回归器, 这些决策树的形成采用了随机的方法, 因此也叫做随机决策树, 决策树之间是没有关联, 当测试数据进入随机森林时, 每棵决策树对数据进行回归, 最终取所有决策树中回归结果最多的一类作为最终输出的结果。将同一年的国内碳排放量及其 7 个主要影响因素数据作为一个样本点, 设样

本的属性个数为 R , $r \in Z^+, 0 < r < R$, 由于本体样本点较少, 随机森林重采样的输出结果质量较差, 故本文采用贝叶斯算法对随机森林的超参数进行优化, 以得到最优的超参数组合, 提高随机森林模型的性能表现

将国内碳排放量作为响应变量, 7 个主要影响因素: 农村碳排放量, 人口密度, 人均可支配收入, 本科专科毕业生数, 能源消费总量, 公共图书馆业机构, 工业企业单位数作为预测变量。

在构建决策树前, 本文首先通过主成分分析 (PCA) 的方法, 对数据进行降维处理, 得出各影响因素的累积贡献率的直方图如图15所示:

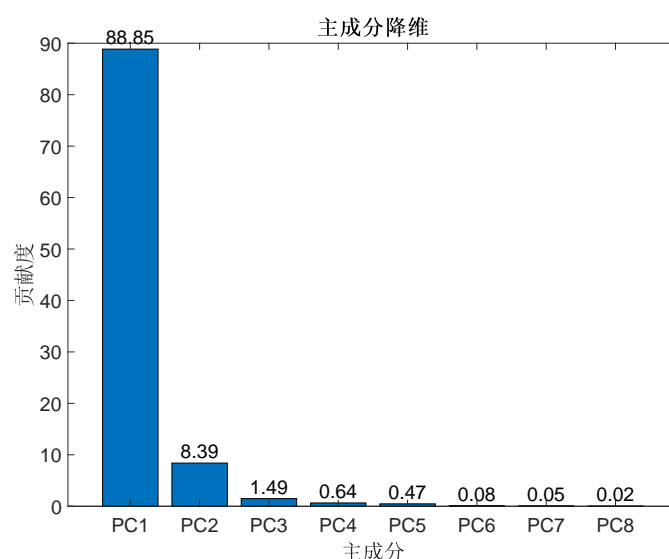


图 15 主成分贡献率直方图

由于 PCA 模型仅保留足以解释 95% 方差的成分, 在训练后, 主成分减少为两个, 由于主成分分析后损失的信息较多, 故本文随机森林模型同时进行 PCA 分析和不进行 PCA 分析两种方法进行回归训练。

7.3.2 随机森林的多元回归模型的预测结果

分别进行有 PCA 和无 PCA 的随机森林回归训练, 计算模型中平均预测误差的平方根 $RMSE$ 、平方 MSE 、绝对值 MAE 和决定系数 R^2 , 得到的两种评估指标数据如下表:

模型	$RMSE$	MSE	R^2	MAE
随机森林 (有 PCA)	0.082822263	0.006859527	0.954972247	0.065432281
随机森林 (无 PCA)	0.083441865	0.006962545	0.954296011	0.070921787

分析对比表可以得知, 无 PCA 的随机森林回归模型的预测性能更好, 拟合效果更

优，预测值和实际值之间的偏差更小。故采用无 PCA 降维的随机森林回归模型，预测国内碳排放量的结果如下图16所示：

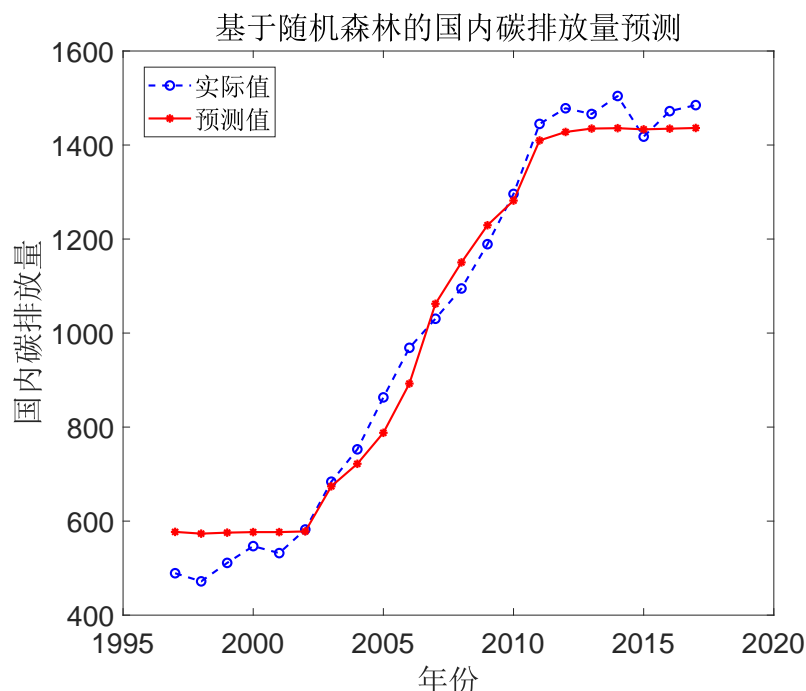


图 16 随机森林预测结果

可见预测值的趋势与实际值的趋势基本吻合。

7.4 基于 LSTM 的时间序列预测

由于预测变量与响应变量都是时间序列，具有明显的趋势性，故本文采用时间序列预测模型预测 7 个预测变量从 2018 到 2032 年的值。首先采用 ARIMA、岭回归、灰色预测模型等统计时间序列模型进行预测，结果发现效果较差。因此本文采用 LSTM 长短时记忆网络模型进行预测。

7.4.1 LSTM 模型

LSTM(Long short-term memory 长短期记忆网络)通过细胞状态和门控机制来管理控制信息的流动，使其能够有效地捕捉时间序列数据中的长期依赖关系，是一种强大的时间序列预测工具。主要优点可以概括为：

- **长期记忆能力：**引入“细胞状态”内部状态，捕捉和保留序列中的长期依赖关系；
- **门控机制：**遗忘门、输入门和输出门这些门控单元通过学习来控制信息的流动，有效地管理细胞状态的更新和选择性遗忘；
- **反向传播稳定性：**由于 LSTM 的门控机制，它在反向传播时通常不会出现梯度消失或爆炸的问题，训练更加容易。

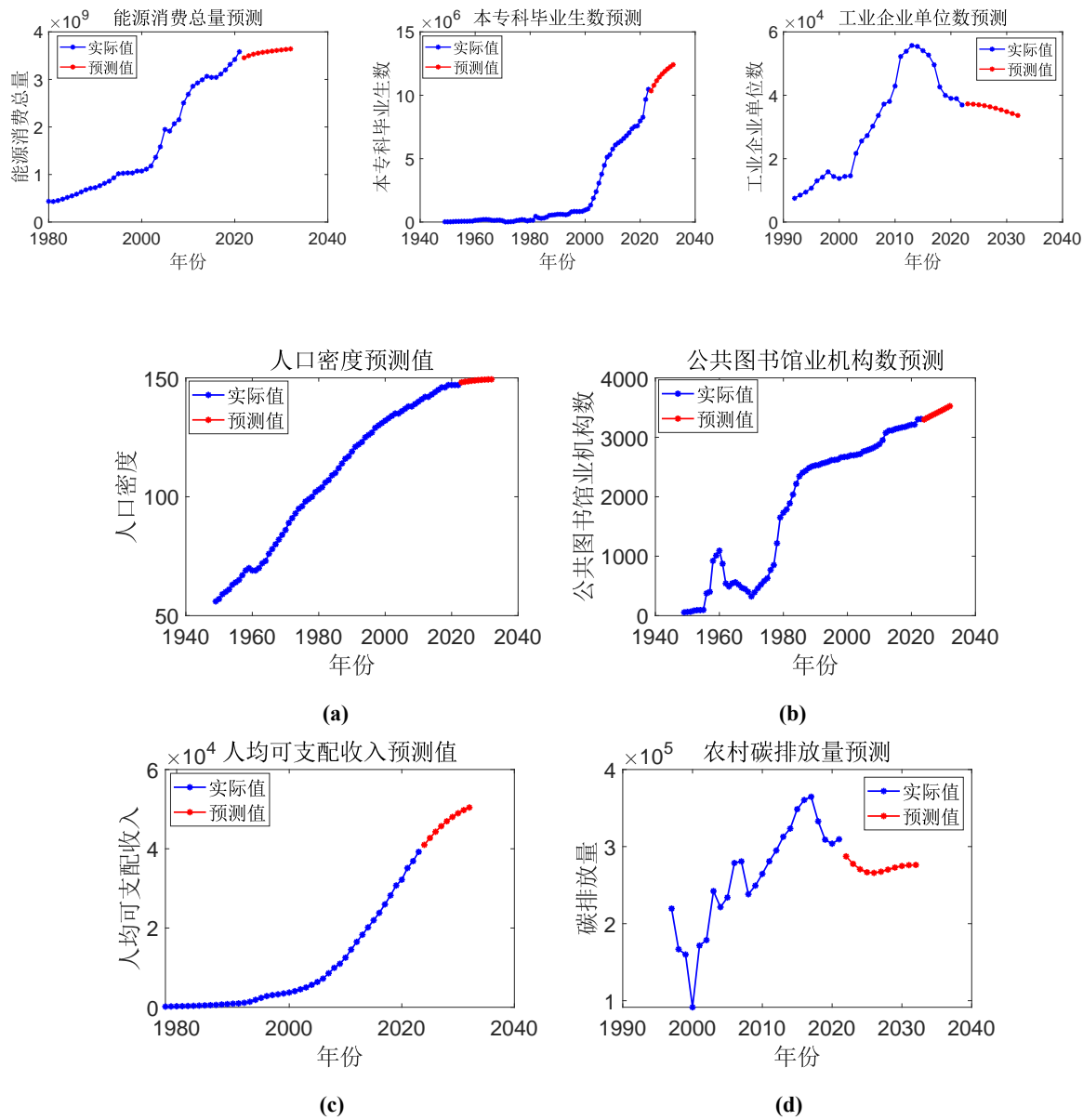
7.4.2 将预测变量代入 LSTM 模型

搭建 LSTM 模型对 7 个预测变量：农村碳排放量，人口密度，人均可支配收入，本
科专科毕业生数，能源消费总量，公共图书馆业机构，工业企业单位数作进行时间序列
预测，LSTM 模型参数如下：

表 3 LSTM 模型参数

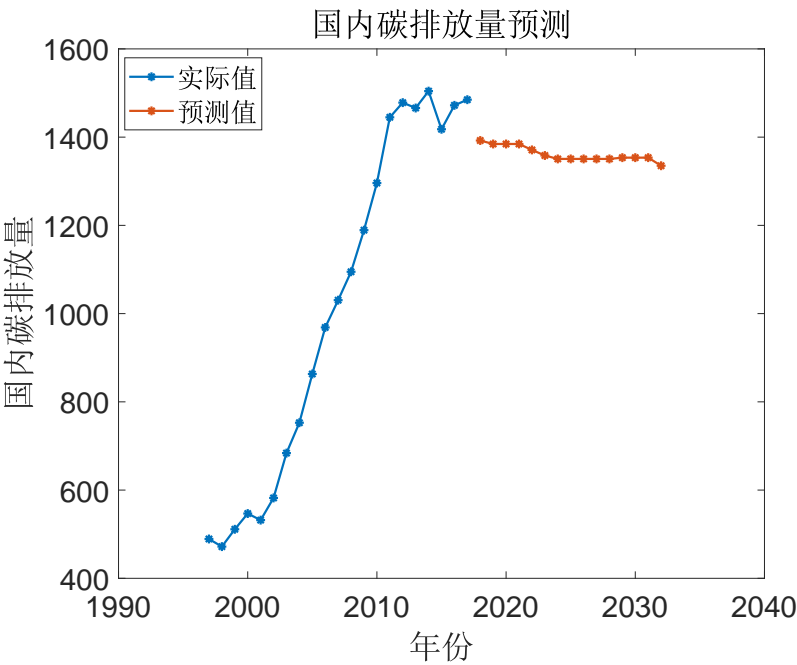
单层 LSTM 单元数	6	损失函数	均方误差
全连接层神经元数	7	优化器	Adam
迭代次数	300	样本批量	7

7 个预测变量的在 2018-2032 年的预测结果如下图所示：



7.5 基于随机森林模型预测未来碳排放量

将 LSTM 预测模型得到各预测变量在 2018-2032 年的预测值代入随机森林的回归模型中，可得国内碳排放量的在 2018-2032 年预测值，国内碳排放的预测图如下：



分析上图可知，国内碳排放量在 2018 年后开始逐渐、缓慢下降，故可推断 2030 年达到碳达峰的目标可行。

八、 问题四的模型的建立和求解

随着全球气候变化和环境问题日益严重，减少碳排放已成为国际社会共同关注的重要议题。人工智能技术作为 21 世纪最具创新性和影响力的技术之一，在科技减碳方面展现出巨大的潜力和价值。本文将从优化能源结构、提高能效、智能交通系统、建筑智能化、精准农业技术、碳捕捉和存储技术（CCS）以及智能数据分析和监测等方面，探讨人工智能技术在科技减碳方面的建议。

8.1 优化能源结构

人工智能技术可以通过数据分析和预测模型，为能源结构的优化提供科学依据。例如，通过对不同能源类型的产能、消耗、碳排放等因素的综合评估，为政策制定者提供决策支持，推动清洁能源如太阳能、风能等的大规模开发和利用。此外，人工智能技术还可以优化能源分配和调度，提高能源利用效率，降低碳排放。

8.2 提高能效

在工业生产和日常生活中，人工智能技术可以通过智能调度、优化生产流程、改进设备运行状态等方式，实现能效的提高。例如，智能制造系统可以收集和分析生产数据，实现智能调度和优化，减少不必要的能源消耗。同时，人工智能技术还可以应用于智能能源管理，实时监控能源消耗情况，优化能源配置，减少能源浪费。

8.3 智能交通系统

智能交通系统是人工智能技术在交通领域的重要应用之一。通过智能交通管控系统，可以实现对交通流量的优化控制，减少交通拥堵，提高交通效率，从而减少交通排放。此外，人工智能技术还可以加速电动汽车充电站的建设与管理，推动电动交通的发展，进一步减少碳排放。

8.4 建筑智能化

在建筑领域，人工智能技术可以通过建筑模拟和能源消耗模型的分析，帮助设计师和规划者优化建筑设计和城市布局，实现更加节能和环保的建筑和城市。同时，人工智能技术还可以通过智能控制系统，实现建筑内部能源的智能管理，降低能源消耗，减少碳排放。

8.5 精准农业技术

精准农业技术通过人工智能驱动的工具提供作物管理、灌溉调度和病虫害防治方面的见解，实现更可持续、更环保的农业实践。人工智能技术可以根据土壤、气候和作物生长情况，制定科学的种植方案，减少化肥和农药的使用，降低农业生产过程中的碳排放。

8.6 碳捕捉和存储技术（CCS）

人工智能技术可以优化碳捕获和储存流程，提高 CCS 技术的效率和可行性。通过机器学习算法，可以分析与 CCS 操作相关的大量数据集，提高在碳排放进入大气之前捕获碳排放的整体效率和可行性。这将有助于实现工业过程中的大规模减碳，为应对气候变化提供重要支持。

8.7 智能数据分析和监测

智能数据分析和监测是实现科技减碳的重要手段之一。通过实施人工智能监测系统，可以持续跟踪和报告各个行业的碳排放情况。实时数据分析和报告机制使组织和政

府能够评估其环境影响，并采取积极措施减少排放。此外，人工智能还可以通过分析历史数据和预测模型，为减碳政策的制定和实施提供科学依据。

8.8 结论：

人工智能技术在科技减碳方面具有广泛的应用前景和巨大的潜力。通过优化能源结构、提高能效、智能交通系统、建筑智能化、精准农业技术、碳捕捉和存储技术（CCS）以及智能数据分析和监测等手段的综合运用，可以有效降低碳排放，推动全球绿色低碳发展。未来，随着人工智能技术的不断发展和完善，其在科技减碳方面的作用将更加显著和重要。

九、模型的评价

9.1 模型的优点

- 优点 1: Pearson 相关性分析简单易懂，表达直观，可解释性强。
- 优点 2: 随机森林对于数据中的噪声和缺失值具有很强的鲁棒性，能够处理复杂的数据集，准确性高，训练速度快，且不容易过拟合。
- 优点 3: LSTM 模型可以用于处理各种类型的时间序列数据，包括具有复杂模式和非线性关系的数据，且通常能够实现较高的预测准确性。

9.2 模型的缺点

- 缺点 1: Pearson 相关性分析只能检测到线性关系，对于非线性关系的相关性无法很好地反映，且要求变量符合正态分布，如果数据不符合正态分布，则相关性分析结果可能不准确。
- 缺点 2: 随机森林是一种黑盒模型，难以解释为什么会做出特定的预测。
- 缺点 3: LSTM 模型在处理小型数据集或数据噪声较大的情况下，可能存在过拟合的风险，需要采取适当的正则化方法。并且 LSTM 模型有许多超参数需要调优，如网络结构、学习速率等，不当的选择可能会影响模型的性能。

参考文献

- [1] 司守奎, 孙玺菁. 数学建模算法与应用[M]. 北京: 国防工业出版社, 2011.
- [2] 卓金武. MATLAB 在数学建模中的应用[M]. 北京: 北京航空航天大学出版社, 2011.
- [3] 白万平, 杨广仁, 张学敏. 碳排放增加与气温变化统计因果关系的多重检验[J]. 贵州财经大学学报, 2013:46-51.
- [4] 高杨, 李滨, 冯振, 等. 全球气候变化与地质灾害响应分析[J]. 地质力学学报, 2017, 23:65-77.
- [5] LAMB W F, WIEDMANN T, PONGRATZ J, et al. A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018[J]. Environmental research letters, 2021, 16(7):073005.
- [6] LASHOF D A, AHUJA D R. Relative contributions of greenhouse gas emissions to global warming[J]. Nature, 1990, 344(6266):529-531.
- [7] WEI T, DONG W, YAN Q, et al. Developed and developing world contributions to climate system change based on carbon dioxide, methane and nitrous oxide emissions[J]. Advances in atmospheric sciences, 2016, 33:632-643.
- [8] REHMAN A, MA H, IRFAN M, et al. Does carbon dioxide, methane, nitrous oxide, and ghg emissions influence the agriculture? evidence from china[J]. Environmental Science and Pollution Research, 2020, 27:28768-28779.
- [9] KWILINSKI A, DOBROVOLSKA O, WOŁOWIEC T, et al. Carbon dioxide, nitrous oxide, and methane: What types of greenhouse gases are most affected by green investments and renewable energy development?[J]. Energies, 2024, 17(4):804.
- [10] 张志强, 曲建升, 曾静静. 温室气体排放评价指标及其定量分析[J]. 地理学报, 2008, 63(7):10.
- [11] 联合国. 联合国气候变化框架公约[J]. 联合国公约, 1992, 1(1):25.
- [12] 联合国. 联合国气候变化框架公约[J]. 联合国公约, 1998, 1(2):23.
- [13] 王慧, 王艳娟, 周春雨. 浅析温室气体与气候变化的影响[J]. 统计与咨询, 2015(3):3.
- [14] 黄炜, 王诚. 温室气体排放总量目标制定研究方法述评[J/OL]. 价值工程, 2017, 36: 246-247. DOI: 10.14018/j.cnki.cn13-1085/n.2017.12.100.

- [15] 郑小慧. 中国区域性碳排放影响因素和碳达峰预测研究[D]. [出版地不详]: 天津财经大学, 2022.
- [16] 佟昕, 陈凯, 李刚, 等. 中国碳排放与影响因素的实证研究——基于 2000~2011 年中国以及 30 个省域的灰色关联分析[J]. 工业技术经济, 2015(3):66-78.
- [17] 芮涛涛. 中国城市 2030 年碳达峰预测及路径研究[D]. [出版地不详]: 陕西师范大学, 2022.
- [18] 刘昌新, 秦小乔, 朱永彬. 人工智能助力碳中和目标实现的基本途径与内在机理[J/OL]. 环境保护, 2023, 51:42-46. DOI: 10.14026/j.cnki.0253-9705.2023.06.009.
- [19] 王豪杰, 花雅雯, 雷万鹏, 等. 人工智能在“碳中和”行业中的应用及展望[J/OL]. 林业机械与木工设备, 2023, 51:21-25+30. DOI: 10.13279/j.cnki.fmwe.2023.0057.
- [20] 王一涵. 全球“碳中和”进程下人工智能技术的应用场景与效应评估[J/OL]. 中国高新科技, 2023:36-38. DOI: 10.13535/j.cnki.10-1507/n.2023.10.11.
- [21] 白舒婕. 人工智能为产业减碳注入“智慧动能”[R]. [出版地不详]: 国际商报, 2023.

附录 A 文件列表

文件名	功能描述
pro1.m	问题一程序代码
pro2.m	问题二程序代码
pro3.m	问题三程序代码
LSTM.m	LSTM 模型代码

附录 B 代码

pro1.m

```
1 %1. 请分析碳排放量、自然灾害发生量与地球地表温度之间的关系
2 disaster = readtable("自然灾害数据.csv");
3 disaster_year = disaster.Year;
4
5 % 使用histogram函数统计各年份出现的次数
6 histogram_data = histogram(disaster_year);
7
8 disaster_count = histogram_data.Values;
9
10 %% 画图
11 data = readtable("pro1_data.xlsx");
12 %0-1标准化
13 normalized_data = normalize(data, 'range');
14
15 year = linspace(2000, 2020, 21);
16
17 Carbon = normalized_data.Carbon;
18 disaster_count = normalized_data.Disaster;
19 Temp = normalized_data.Temp;
20
21 figure
22
23
24 plot(year, Carbon, 'b-*', 'LineWidth', 1.5)
```

```

25 hold on
26 plot(year,disaster_count,'g-*','LineWidth',1.5)
27 plot(year,Temp,'r-*','LineWidth',1.5)
28 xlabel('年份')
29 ylabel('归一化后的数据')
30 title('碳排放量,自然灾害发生量和地球地表温度的面板数据变化趋势
    图')
31 legend('碳排放量','自然灾害发生量','地球地表温度')
32 set(gca,'FontSize',20)
33
34 %% 增量画图
35 diff_data = diff(table2array(normalized_data));
36 figure
37 plot(year(2:end),diff_data(:,1),'b-*','LineWidth',1.5)
38 hold on
39 plot(year(2:end),diff_data(:,2),'g-*','LineWidth',1.5)
40 plot(year(2:end),diff_data(:,3),'r-*','LineWidth',1.5)
41 xlabel('年份')
42 ylabel('归一化后的数据')
43 title('碳排放量,自然灾害发生量和地球地表温度的增量数据变化趋势
    图')
44 legend('碳排放变化量','自然灾害变化量','地球地表温度变化量')
45 set(gca,'FontSize',20)
46
47 %% 增量相关性分析
48 [corr_diff,pval1] = corr(diff_data) ;
49 figure
50
51 % 相关性矩阵热图
52 h_diff = heatmap(normalized_data.Properties.VariableNames, ...
53     normalized_data.Properties.VariableNames,corr_diff);
54 title('增量间的相关性热力图')
55 set(gca, 'FontSize', 20);
56
57 %% 相关性分析

```



```

58 [correlationMatrix,pval] = corr(table2array(normalized_data))
59 figure
60
61 % 相关性矩阵热图
62 h = heatmap(normalized_data.Properties.VariableNames, ...
63     normalized_data.Properties.VariableNames,correlationMatrix
64     );
65 title('面板数据间的相关性热力图')
66 set(gca, 'FontSize', 20);
67 %% 二元回归
68 y = Temp;
69 X = [ones(size(Carbon)) Carbon disaster_count Carbon.*
70     disaster_count];
71 [b,bint,r,rint,stats] = regress(y,X)
72 figure
73 scatter3(Carbon,disaster_count,y,'filled')
74 hold on
75 Carbonfit = min(Carbon):0.01:max(Carbon);
76 disaster_countfit = min(disaster_count):0.01:max(
77     disaster_count);
78 [CarbonFIT,disaster_countFIT] = meshgrid(Carbonfit,
79     disaster_countfit);
80 YFIT = b(1) + b(2)*CarbonFIT + b(3)*disaster_countFIT + b(4)*
81     CarbonFIT.*disaster_countFIT;
82 mesh(CarbonFIT,disaster_countFIT,YFIT)
83 xlabel('Carbon')
84 ylabel('Disaster')
85 zlabel('Temp')
86 view(50,10)
87 hold off
88 legend('data','多重回归拟合曲线')
89 set(gca, 'FontSize', 20);
90 %% 3个一元回归
91 OLS = zeros(3,2);
92 figure

```

```

88  sgtitle('最小二乘拟合', 'FontWeight', 'bold', 'FontSize', 20)
89  for i=1:2
90  if i ==1
91      x = Carbon;
92      y = disaster_count;
93      v1 = {'碳排放量'};
94      v2 = {'自然灾害发生量'};
95  end
96  if i ==2
97      x = Carbon;
98      y = Temp;
99      v1 = {'碳排放量'};
100     v2 = {'地球地表温度'};
101  end
102  if i ==3
103      x = Temp;
104      y = disaster_count;
105      v1 = {'地球地表温度'};
106      v2 = {'自然灾害发生量'};
107  end
108  %figure
109  fitresult = fit(x, y, 'poly1' );
110  % OLS(i,1) = fitresult.p1;
111  % OLS(i,2) = fitresult.p2;
112  [p,S] = polyfit(x,y,1);
113  [y_fit,delta] = polyval(p,x,S);
114
115  subplot(1,2,i)
116  plot(x,y,'bo','LineWidth',1.5)
117  hold on
118  plot(x,y_fit,'r-','LineWidth',1.5)
119  plot(x,y_fit+2*delta,'m--',x,y_fit-2*delta,'m--','LineWidth'
    ,1.5)
120  title('Linear Fit of Data with 95% Prediction Interval')
121  legend('Data','Linear Fit','95% Prediction Interval')

```

```

122 % myplot = plot(fitresult,x,y);
123 xlabel(v1)
124 ylabel(v2)
125 title(['y = ',num2str(fitresult.p1),' x + ',num2str(fitresult.
    p2)]]
126 %set(myplot,'linewidth',1.5,'MarkerSize',18)
127 set(gca, 'FontSize', 15);
128 end
129
130 %%

```

pro2.m

```

1 %% 数据读取
2 data_2 = readtable("pro2.xlsx",'Sheet','Sheet1');
3 raw3 = readtable("pro2.xlsx",'Sheet','com3');
4
5 %0-1标准化
6 normalized_data_3 = normalize(raw3, 'range');
7
8 normalized_data_2 = normalize(data_2, 'range');
9 NO = normalized_data_2.NO;
10 CH4 = normalized_data_2.CH4;
11 C = normalized_data_2.C;
12 Global = normalized_data_2.Global;
13
14 year2 = linspace(2010, 2020, 11);
15
16 year3 = linspace(1990, 2020, 31);
17
18 %% 趋势图 3 before 归一化
19 figure
20 plot(year3,raw3.NO,'r-*','LineWidth',1.5)
21 hold on
22 plot(year3,raw3.CH4,'g-*','LineWidth',1.5)
23 plot(year3,raw3.C,'b-*','LineWidth',1.5)

```

```

24 %plot(year2,Global,'r-o','LineWidth',1.5)
25 xlabel('年份')
26 ylabel('排放量（千公吨二氧化碳当量）')
27 title('一氧化氮、甲烷、碳排放量的变化趋势图')
28 legend('一氧化氮排放量','甲烷排放量','碳排放量')
29 set(gca,'FontSize',20)
30
31
32 %% 趋势图 3 after 归一化
33 figure
34 plot(year3,normalized_data_3.NO,'r-*','LineWidth',1.5)
35 hold on
36 plot(year3,normalized_data_3.CH4,'g-*','LineWidth',1.5)
37 plot(year3,normalized_data_3.C,'b-*','LineWidth',1.5)
38 %plot(year2,Global,'r-o','LineWidth',1.5)
39 xlabel('年份')
40 ylabel('归一化后的数据')
41 title('一氧化氮、甲烷、碳排放量的变化趋势图')
42 legend('一氧化氮排放量','甲烷排放量','碳排放量')
43 set(gca,'FontSize',20)
44
45 %% 趋势图 4
46 figure
47 plot(year2,NO,'-*',"Color",[0.9290 0.6940 0.1250],'LineWidth',1.5)
48 hold on
49 plot(year2,CH4,'g-*','LineWidth',1.5)
50 plot(year2,C,'b-*','LineWidth',1.5)
51 plot(year2,Global,'r-o','LineWidth',1.5)
52 xlabel('年份')
53 ylabel('归一化后的数据')
54 title('一氧化氮、甲烷、碳排放量与全球温室气体排放量的变化趋势图')
55 legend('一氧化氮排放量','甲烷排放量','碳排放量','全球温室气体排放量')

```

```

56 set(gca,'FontSize',20)
57
58
59 %% 相关性分析
60 [corr2,pval_world] = corr(table2array(normalized_data_2)) ;
61 figure
62
63 % 相关性矩阵热图
64 h2 = heatmap(normalized_data_2.Properties.VariableNames, ...
65     normalized_data_2.Properties.VariableNames,corr2);
66 title('面板数据间的相关性热力图')
67 set(gca, 'FontSize', 20);

```

pro3.m

```

1  %3. 分析影响国内碳排放变化的主要影响因素和贡献率，建立合适的碳
   排放量模型并预测2030年碳达峰目标实现的可行性
2  %% 数据读取
3  data_China = readtable("pro3.xlsx",'Sheet','Sheet1');
4  %0-1标准化
5  normalized_data_China = normalize(data_China(:,2:end), 'range'
   );
6
7  year_China = data_China.year;
8
9  %% 趋势图
10 figure
11 plot(year_China,normalized_data_China.China_Carbon,'r-*','
   linewidth',2);
12 hold on;
13 plot(year_China,normalized_data_China.Countryside_Carbon,'
   linewidth',1.5);
14 plot(year_China,normalized_data_China.population_density,'m-',
   'linewidth',1.5);
15 plot(year_China,normalized_data_China.income,'linewidth',1.5);
16 plot(year_China,normalized_data_China.graduate,'g-','linewidth

```

```

    ',1.5);
17 plot(year_China,normalized_data_China.energy,'linewidth',1.5);
18 plot(year_China,normalized_data_China.library,'linewidth',1.5)
    ;
19 plot(year_China,normalized_data_China.SO2,'linewidth',1.5);
20 plot(year_China,normalized_data_China.enterprise,'linewidth'
    ,1.5);
21 %set(China_plot,'linewidth',1.5)
22 xlim([1997,2022]);
23 legend('国内碳排放量','农村碳排放量','人口密度','人均可支配收
    入','本专科毕业生数','能源消费总量','公共图书馆业机构数','国
    内二氧化硫排放量','工业企业单位数')
24 xlabel('年份')
25 ylabel('归一化后的数据')
26 title('国内碳排放与各影响因素的变化趋势图')
27 set(gca,'FontSize',20)
28 hold off;
29
30
31 %% 相关性分析
32
33 variable = {'国内碳排放量','农村碳排放量','人口密度',...
34     '人均可支配收入','本专科毕业生数','能源消费总量',...
35     '公共图书馆业机构数','国内二氧化硫排放量','工业企业单位数'
    };
36 [corr_China,pval] = corr(table2array(normalized_data_China))
37 figure
38
39 % 相关性矩阵热图
40 h = heatmap(variable, ...
41     variable,corr_China);
42 title('国内碳排放与各影响因素的相关性热力图')
43 set(gca, 'FontSize', 20);
44 %% 相关性排序
45 R = corr_China(1,2:end);

```

```

46 factors_lable = variable(2:end);
47 % 将数组 R 转换为表，列的名称是1到22
48 %T = array2table(R, 'VariableNames',cellstr(num2str((1:22))))
    ;
49
50 % 对T表按照第一行降序排序
51 [~, idx] = sort(abs(R), 'descend');
52 T_sorted = R(idx);
53 L_sorted = factors_lable(idx);
54 figure
55 bar(T_sorted);
56 xticks(1:8);
57 xticklabels(L_sorted);
58 xlabel('影响因素');
59 ylabel('相关系数');
60 title('中国碳排放量与影响因素之间的相关系数', 'FontWeight', '
    bold', 'FontSize', 20);
61 % 在条形图上显示数值
62 for i = 1:length(T_sorted)
63     text(i, T_sorted(i), num2str(T_sorted(i), '%.3f'), '
        HorizontalAlignment', ...
64         'center', 'VerticalAlignment', 'bottom','FontSize',23)
    ;
65 end
66 set(gca,'FontSize',20)
67
68 %% 主成分分析
69 factors = table2array(normalized_data_China);
70 factors(:,2) = [];
71 %%
72 [coeff,score,latent,tsquared,explained] = pca(factors)
73 num = 8;
74 X = categorical({'PC1','PC2','PC3','PC4','PC5','PC6','PC7','
    PC8'});
75 explained_pc = explained(1:num,1);

```

```

76 contribution = sum(explained_pc);
77 for i = 1:num
78     fprintf('第%d个主成分的贡献度为%f%%\n',i,explained_pc(i,1)
79         );
80     fprintf('主成分总贡献度为%f%%\n',contribution);
81
82     figure
83     bar(X,explained_pc);%variable(2:end),
84     xlabel('主成分');
85     ylabel('贡献度');
86     title('主成分降维', 'FontWeight', 'bold', 'FontSize', 14);
87     for i = 1:length(explained_pc)
88         text(i, explained_pc(i), num2str(explained_pc(i), '%.2f'),
89             'HorizontalAlignment', ...
90             'center', 'VerticalAlignment', 'bottom','FontSize',20)
91         ;
92     end
93     set(gca,'FontSize',20)
94
95 %% regress model
96 regress_train_data = readtable("pro3.xlsx",'Sheet','predict2')
97 ;
98
99 %% time series
100 countryside =readtable("pro3.xlsx",'Sheet','countryside');
101 T_countryside = countryside.countryside;
102 % 将数据存储在变量中
103 year = [1997:2021];
104
105 % 创建时间序列对象
106 ts = timeseries(T_countryside, year);
107
108 % 拟合ARIMA模型
109 mdl = arima(2,5,2); % 定义ARIMA模型的参数

```



```

107 % 将时间序列数据拟合到ARIMA模型中
108 fit = estimate mdl, ts.Data);
109
110 % 预测未来的值
111 futureYears = [2022:2035];
112 [countrysidePredicted, ~] = forecast(fit, length(futureYears))
    ;
113
114 % 可视化结果
115 figure;
116 plot(year, T_countryside, 'b.-', 'DisplayName', 'Actual Data')
    ;
117 hold on;
118 plot(futureYears, countrysidePredicted, 'r.-', 'DisplayName',
    'Predicted Data');
119 xlabel('Year');
120 ylabel('Countryside Carbon Emission');
121 legend;
122 title('Prediction of Countryside Carbon Emission');

```

LSTM.m

```

1 %variable = {'国内碳排放量','农村碳排放量','人口密度',...
2     % '人均可支配收入','本专科毕业生数','能源消费总量',...
3     % '公共图书馆业机构数','国内二氧化硫排放量','工业企业单位
    数'};
4 %% 农村碳排放量
5 T = readtable("LSTM_pre1.xlsx");
6 countryside_actual = T.countryside;
7 %countryside_predict = Lstm_Forecast_Value;
8 year_countryside = linspace(1997, 2021, 25);
9 year_countryside1 = linspace(2022, 2032, 11);
10 figure
11 plot(year_countryside,countryside_actual,'b-*','LineWidth'
    ,1.5)
12 hold on

```

```

13 plot(year_countryside1,countryside_predict,'r-*','LineWidth'
    ,1.5)
14 xlabel('年份')
15 ylabel('碳排放量')
16 title('农村碳排放量预测')
17 legend('实际值','预测值')
18 set(gca,'FontSize',20)
19
20 %% 人口密度
21 T = readtable("LSTM_pre2.xlsx");
22 density_actual = T.population_density;
23 density_predict = Lstm_Forecast_Value;
24 year_density = linspace(1949, 2022,74);
25 year_density1 = linspace(2023, 2032, 10);
26 figure
27 plot(year_density,density_actual,'b-*','LineWidth',1.5)
28 hold on
29 plot(year_density1,density_predict,'r-*','LineWidth',1.5)
30 xlabel('年份')
31 ylabel('人口密度')
32 title('人口密度预测')
33 legend('实际值','预测值')
34 set(gca,'FontSize',20)
35
36
37 %% 收入
38 T = readtable("LSTM_pre3.xlsx");
39 income_actual = T.income;
40 income_predict = Lstm_Forecast_Value;
41 year_income = linspace(1978, 2023, 46);
42 year_income1 = linspace(2024, 2032, 9);
43 figure
44 plot(year_income,income_actual,'b-*','LineWidth',1.5)
45 hold on
46 plot(year_income1,income_predict,'r-*','LineWidth',1.5)

```

```

47 xlabel('年份')
48 ylabel('人均可支配收入')
49 title('人均可支配收入预测')
50 legend('实际值','预测值')
51 set(gca,'FontSize',20)
52
53 %% 毕业生
54 T = readtable("LSTM_pre4.xlsx");
55 graduate_actual = T.graduate;
56 graduate_predict = Lstm_Forecast_Value;
57 year_graduate = linspace(1949, 2023, 75);
58 year_graduate1 = linspace(2024, 2032, 9);
59 figure
60 plot(year_graduate,graduate_actual,'b-*','LineWidth',1.5)
61 hold on
62 plot(year_graduate1,graduate_predict,'r-*','LineWidth',1.5)
63 xlabel('年份')
64 ylabel('本专科毕业生数')
65 title('本专科毕业生数预测')
66 legend('实际值','预测值')
67 set(gca,'FontSize',20)
68
69 %% 能源
70 T = readtable("LSTM_pre5.xlsx");
71
72 energy_actual = T.energy;
73 energy_predict = Lstm_Forecast_Value;
74 year_energy = linspace(1980, 2021, 42);
75 year_energy1 = linspace(2022, 2032, 11);
76 figure
77 plot(year_energy,energy_actual,'b-*','LineWidth',1.5)
78 hold on
79 plot(year_energy1,energy_predict,'r-*','LineWidth',1.5)
80 xlabel('年份')
81 ylabel('能源消费总量')

```

```

82 title('能源消费总量预测')
83 legend('实际值','预测值')
84 set(gca,'FontSize',20)
85
86
87 %% 图书馆
88 T = readtable("LSTM_pre6.xlsx");
89
90 lib_actual = T.lib;
91 lib_predict = Lstm_Forecast_Value;
92 year_lib = linspace(1949, 2023, 75);
93 year_lib1 = linspace(2024, 2032, 9);
94 figure
95 plot(year_lib,lib_actual,'b-*','LineWidth',1.5)
96 hold on
97 plot(year_lib1,lib_predict,'r-*','LineWidth',1.5)
98 xlabel('年份')
99 ylabel('公共图书馆业机构数')
100 title('公共图书馆业机构数预测')
101 legend('实际值','预测值')
102 set(gca,'FontSize',20)
103
104
105 %% 企业
106 T = readtable("LSTM_pre7.xlsx");
107
108 enterprise_actual = T.enterprise;
109 enterprise_predict = Lstm_Forecast_Value;
110 year_enterprise = linspace(1992, 2022, 31);
111 year_enterprise1 = linspace(2023, 2032, 10);
112 figure
113 plot(year_enterprise,enterprise_actual,'b-*','LineWidth',1.5)
114 hold on
115 plot(year_enterprise1,enterprise_predict,'r-*','LineWidth'
    ,1.5)

```

```

116 xlabel('年份')
117 ylabel('工业企业单位数')
118 title('工业企业单位数预测')
119 legend('实际值','预测值')
120 set(gca,'FontSize',20)
121
122 %% 预测模型检验
123 regress_train_data = readtable("pro3.xlsx",'Sheet','predict2')
124     ;
125 China_C_actual = regress_train_data.China_Carbon;
126 regress_train_variable = regress_train_data(:,1:7);
127 China_C_regress_predict = rf_trainedModel.predictFcn(
128     regress_train_variable);
129 %year_China1 = linspace(2018, 2032, 15);
130 figure
131 hold on
132 plot(year_China,China_C_actual,'b--o','LineWidth',1.5)
133 plot(year_China,China_C_regress_predict,'r-*','LineWidth',1.5)
134 xlabel('年份')
135 ylabel('国内碳排放量')
136 title('基于随机森林的国内碳排放量预测')
137 box on
138 legend('实际值','预测值')
139 set(gca,'FontSize',20)
140 hold off
141
142 %% 预测2018-2032的中国碳排放量
143 V_predict = readtable("pro3.xlsx",'Sheet','predict4');
144 China_C_actual = data_China.China_Carbon;
145 China_C_future_predict = rf_trainedModel.predictFcn(V_predict
146     (:,1:end-1));
147 China_predict = [China_C_regress_predict;
148     China_C_future_predict];

```

```
147
148 year_China1 = linspace(2018, 2032, 15);
149 year_China2 = linspace(1997, 2032, 36);
150 figure
151 hold on
152 plot(year_China,China_C_actual,'-*','LineWidth',1.5)
153 plot(year_China1,China_C_future_predict,'-*','LineWidth',1.5)
154 %plot(year_China2,China_predict,'--','Color',[0.8500 0.3250
    0.0980],'LineWidth',1.5)
155 box on
156 xlabel('年份')
157 ylabel('国内碳排放量')
158 title('国内碳排放量预测')
159 legend('实际值','预测值')
160 set(gca,'FontSize',20)
161 hold off
```