

目录

摘要.....	1
一. 研究背景.....	2
二. 相关研究与研究创新点.....	3
三. 数据的来源与分析.....	6
3.1 数据来源	6
3.1.1 各省碳排放数据与中国碳排放年度数据:	6
3.1.2 碳排放数据中各气体的含量数据:	6
3.1.3 社会经济指标数据:	6
3.2 数据分析	6
3.2.1 各省碳排放数据与中国碳排放年度数据分析	6
3.2.2 碳排放数据中各气体的含量数据分析:	9
3.2.3 社会经济指标指标数据分析	10
四. 模型的建立与求解.....	11
4.1 碳排放物间的关联分析与主成分分析	11
4.1.1 碳排放物的主成分分析	11
4.1.2 碳排放物间的关联分析	15
4.2 碳排放预测模型的建立	16
4.2.1 模型介绍	16
4.2.2 平稳性检验	16
4.2.3 模型拟合确定最优 p, q 值	17
4.2.4 模型的检验	19
4.2.5 碳排放量预测和结果分析	21
五. 结论.....	24
5.1 结论一	24
5.2 结论二	24
六. 建议与启示.....	25
6.1 绿色经济发展策略	25
6.2 能源结构优化策略	25

6.3 工业生产领域技术革新策略25

6.4 燃料使用优化战略26

七. 附录.....26

八. 参考文献.....35

致谢.....36

表格与插图清单

图 1 近 25 年总和燃料使用情况3

图 2 总体流程图5

图 3 总年份碳排放数据箱型图7

图 4 年份处理后的碳排放箱型图7

图 5 碳排放量前五的省份年度数据8

图 6 各省碳排放数据9

图 7 不同种碳气体的年份箱型图9

图 8 化合物占比饼状图10

图 9 GDP 与碳排放数据的拟合曲线.....10

图 10 辅助变量热力图11

图 11 KMO 检验结果13

图 12 主成分分析结果一14

图 13 主成分分析结果二14

图 14 主成分分析流程图14

图 15 因素间的灰色关联度分析15

图 16 ADF 检验结果17

表 1 样本自相关系数与偏相关系数17

图 17 BIC 与 AIC 检验结果.....18

图 18 模型运行结果19

图 19 残差检验双 Q 图.....20

图 20 ARIMA 流程图20

图 21 实际与预测对比图21

图 22 碳排放量的整合21

图 23 碳排放的预测值22

图 24 原煤排放量预测图22

图 25 原煤排放量预测22

图 26 焦炭排放量预测图23

图 27 焦炭排放量预测23

图 28 柴油机油排放量预测图23

图 29 柴油机油排放量预测24

摘要

能源消费是碳排放的主要来源，而我国正处于高速发展的阶段，碳排放量的增长是大势所趋，但绿水青山就是金山银山，并且我国积极响映国际号召，预计在 2030 年达到碳排放量的峰值，但我国目前的减碳目标仍面临着巨大的挑战。为了推进减碳的推进，本文运用了**智能化的 ARIMA（时间序列预测模型）**，灰色关联度分析和主成分分析法预测了未来的碳排放量增长趋势，影响碳排放量的三种主要成分以及影响碳排放量的物质间的相互关系。

本文选择引入国内生产总值，人均生产总值，年末人口数，城镇人数这几种常用的社会经济指标来做为辅助变量来探究影响碳排放量的因素，为了更好的揭示辅助变量与碳排放总量之间的关系，我们引入了**灰色关联系数分析法**得到了相关性热力图并得到了碳排放总量与国内生产总值，人均生产总值，年末人口数具有较强的相关性的结论。同时我们发现碳排放总量与国内生产总值的相关性最强，本文通过进一步研究发现碳排放总量与国内生产总值呈线性关系。

本文中引入了**灰色关联度分析与主成分分析法**来探究碳排放量的物质的主成分与这些物之间的相关性，通过研究我们发现原煤，焦炭和柴油机油这三中碳排放物为主成分，这告诉我们在关注碳排放量时可以重点监测这三种物质的总量与变化趋势。同时通过研究我们给出主要碳排放物间的相关系数热力图，这可以帮助我们制定的政策更加精确有效，使政策具有更多的针对性,同时通过主成分分析法选出的主成分和后文中建立的**智能化的 ARIMA（时间序列预测模型）**来预测主要碳排放物的排放量。

本文建立了**智能化的 ARIMA（时间序列预测模型）**来预测未来的碳排放量以及主要碳排放物的排放量，该模型可以在预测的过程中动态调整差分次数提升准确性。通过研究我们得到了碳排放总量和几种影响碳排放量物质的预测曲线。

关键词：ARIMA（时间序列预测模型） 灰色关联度分析 主成分分析 GDP

一. 研究背景

在当今世界，环境问题已经成为了全球关注的焦点，尤其是碳排放问题。随着工业化、城市化和交通运输的快速发展，人类活动持续释放大量温室气体，进一步加剧了全球气候变化的影响。中国在 2021 年承诺在 2030 年前实现碳达峰，并在 2060 年前实现碳中和，这一宏伟目标将推动中国在减排方面的努力。中国积极参与全球气候变化谈判，展现了其对环保事业的坚定承诺和贡献。早在 2014 年，中美两国就共同发表了《中美气候变化联合声明》，旨在约定于 2023 年左右使二氧化碳排放达到峰值，并努力争取早日实现。中国已逐步成为全球气候治理的引领者，其相关政策彰显了中国人民对“绿色、低碳、环保”社会的远见愿景，突显了大国的责任担当。全球各国在 2016 年签署了《巴黎协定》，旨在协调全球应对气候变化行动，提出了尽快实现温室气体排放达到峰值的共同目标。各国正在努力应对碳排放问题，以减缓气候变化对地球环境的影响，推动可持续发展的进程，这突显了环保事业的至关重要性，需要全球共同努力以降低碳排放，保护地球生态环境的可持续发展。

在过去的 25 年里，作为全球人口众多的大国，中国对能源的需求一直呈持续增长之势。尽管新型清洁能源逐渐进入市场，传统能源的使用逐步减少，但受制于高成本和制造难度等问题，新型能源尚未完全普及。自 1997 年至 2021 年间，原煤的使用量一直保持稳定上升趋势，仍占据主导地位，占比高达 58.49%。天然气位居第二，占比为 13.54%。中国的能源结构正在逐渐从以原煤为主导转向多能源共同发展。然而，原煤的碳排放含量高，对环境影响较大。据数据显示，2014 年中国的碳排放总量位居全球首位，占全球总排放量的 26%，这预示着中国未来将面临更大的减排压力。尽管如此，中国正积极采取措施，推动清洁能源技术的研发与应用，加速能源结构转型，以应对环境挑战和碳排放压力。这一努力不仅是对国内环境可持续发展的承诺，也是对全球环境保护责任的践行，为实现更绿色、更可持续的未来作出了积极贡献。

在一些不足。例如，魏一鸣等人在《碳中和目标下中国碳排放路径研究》[3]中自主构建了中国气候变化综合评估模型/国家能源技术模型（C3IAM/NET），提出了中国碳达峰碳中和路径优化方法，研究了不同经济增速和减排力度情景下的碳排放变化。该研究虽然提供了深入的分析和有价值的见解，但因为它需要大量的假设和参数设定，这可能在短期内提供有效的预测结果，但在长期预测中可能存在较大的不确定性。同样，张帆等人在其研究《共享社会经济路径下中国 2020—2100 年碳排放预测》[4]中，采用了 Cobb-Douglas 生产函数作为经济预测的基础工具，预测了中国在不同 SSP 情景下的社会发展和碳排放情况。尽管该研究提供了有价值的视角，但模型的简化可能未能充分捕捉碳排放与经济、社会、技术等因素之间的复杂相互作用。此外，李永明和张明（2021）研究了碳达峰、碳中和背景下江苏工业面临的挑战、机遇及对策。像这个研究一样，如今许多研究过于专注于特定地区的情况，如江苏省，这可能无法全面反映全国的碳排放和减排情况。

综上所述，尽管现有研究为我们理解碳排放提供了宝贵的信息和策略，但在模型构建、参数设定以及区域代表性方面仍有待进一步完善。未来的研究需要在确保模型准确性的同时，考虑到更多的影响因素，提高模型的适应性和预测的长期准确性，以便更好地指导全国乃至全球的碳减排行动。随着政策的不断推进，在大数据与人工智能时代下，机器学习技术的发展为碳排放的精准监测和有效控制提供了新的途径。

在后疫情时代，研究者刘琰通过构建 ARIMA 模型[8]，并结合指数平滑模型，对中国货物进出口总额进出口数据进行分析，以预测未来的贸易趋势。该研究不仅考虑了时间序列数据的自相关性，还纳入了可能影响贸易的其他变量，如全球经济形势、国内外政策变动等。通过对模型参数的精确估计和模型诊断，研究者能够较为准确地预测未来的进出口总额，并根据预测结果提出针对性的政策建议，如优化贸易结构、增强供应链韧性等，以促进贸易稳定增长和经济健康发展。马晓龙在其研究《旅游业发展影响因子灰色关联分析》[7]中采用灰色关联分析方法来探讨影响旅游业发展的因素。研究确定了旅游业发展的参考指标——旅游总收入，并选取了若干潜在的影响因素作为比较序列，如教育水平、旅游服务设施数量等。通过计算每个因素序列与参考序列之间的关联度，分析它们与旅游总

收入的相似程度。最终，通过比较不同因素的关联度，可以确定它们对旅游业发展的相对重要性。这一研究过程不仅为理解旅游业发展的复杂性提供了定量分析手段，也为政策制定者提供了决策支持。胡孔玉和王竞钊学者利用 PCA 对城市化水平演化进行了分析[6]，从多个维度量化城市化进程，通过提取关键主成分来揭示城市化发展的关键驱动因素和趋势。研究涉及经济、人口、基础设施等指标，通过计算得分和排名来比较不同地区的城市化水平，并为政策制定提供依据，探讨了城市化进程中的关键因素。

鉴于上述研究的多样性和深度，本文通过结合 ARIMA 模型、灰色关联度分析和主成分分析，形成了一个多模型分析框架。这种方法不仅能够从不同角度捕捉碳排放数据的特征，还能够互相验证和补充，提高整体分析的可靠性，为碳排放研究提供了更为全面的视野。

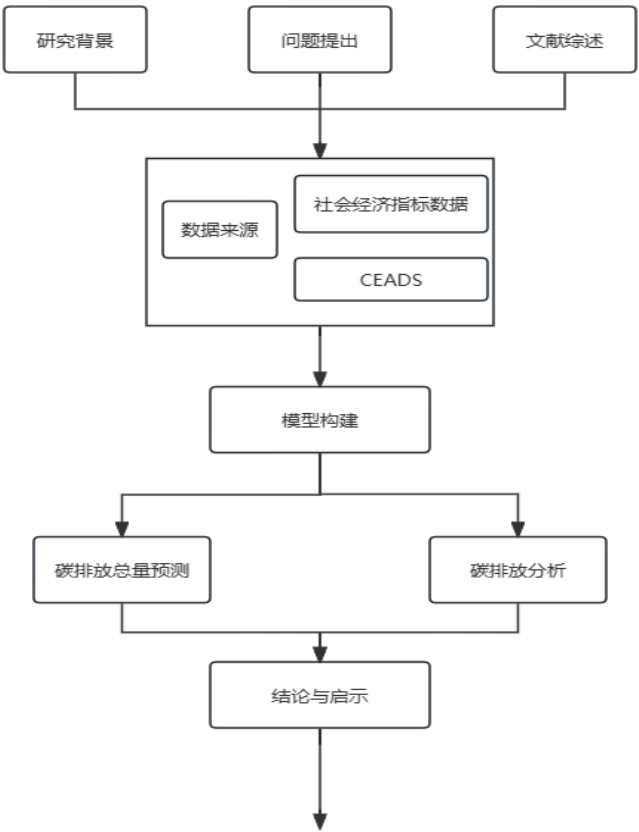


图 2 总体流程图

三. 数据的来源与分析

3.1 数据来源

3.1.1 各省碳排放数据与中国碳排放年度数据：

本文中使用的各省碳排放数据来源于中国碳核算数据库（CEADS）的《1997-2019 年 290 个中国城市碳排放数据》，并且对各个省份的不同城市的每年数据进行加和，得到各省份在不同年份下碳排放数据在对这些数据进行加和得到 1997-2019 各年份碳排放数据总和，为了确保后续的研究结果的准确性，我们通过国家统计局官方网站，《中国统计年鉴》以及各省的统计数据与之进行对照和补充得到了各省的年度碳排放数据与中国碳排放数据。

3.1.2 碳排放数据中各气体的含量数据：

本文中使用的碳排放数据中各气体的含量数据来源于中国碳核算数据库（CEADS）的中国分部门核算碳排放清单 1997-2021》，由于数据中存在异常值与空缺值，我们对数据进行了清洗与插值填补来进行下文中的分析。

3.1.3 社会经济指标数据：

本文选取了一些代表性较强的社会经济指标作为辅助变量来评估碳排放量，本文中的社会经济指标均来自国家统计局官方网站。综合了以往的文献和以往的研究，我们选取了国内生产总值，人均生产总值，年末人口数，城镇人数作为经济指标。

3.2 数据分析

3.2.1 各省碳排放数据与中国碳排放年度数据分析

由于本文中找到的数据是从 1997 到 2019 年的数据，年份跨度较长，部分年份的数据缺失较多，我们对这些数据进行了初步的处理与清洗得到了下图：

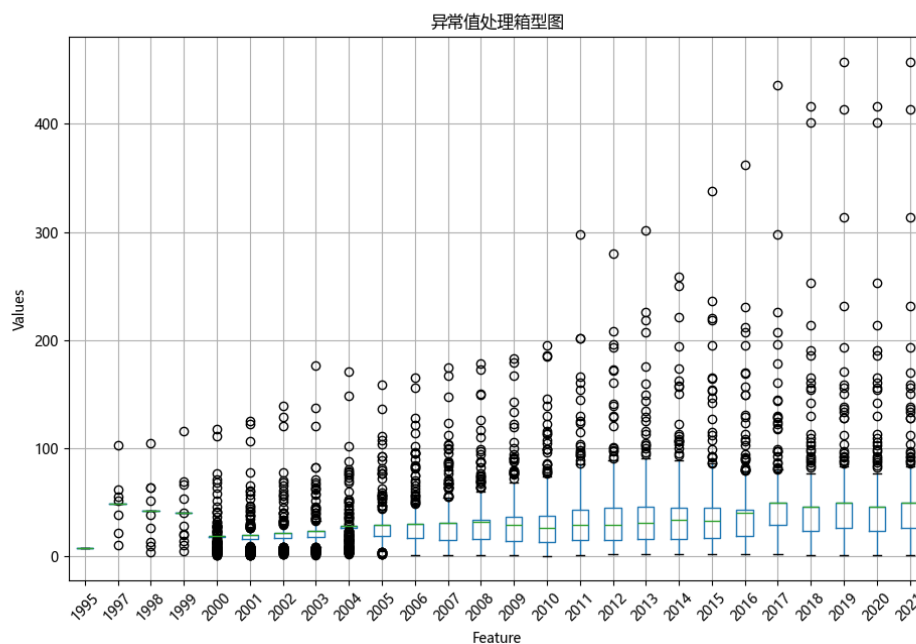


图 3 总年份碳排放数据箱型图

通过图片我们可以看到该图片中的离群点较多, 如果对其进行分析会导致结果的异常与分析的错误, 为了使结果的准确性更高, 本文重新对数据进行了处理。

选取年份从 2000 年到 2021 年对数据进行重新的清洗和插值填补。得到了下面的箱型图:

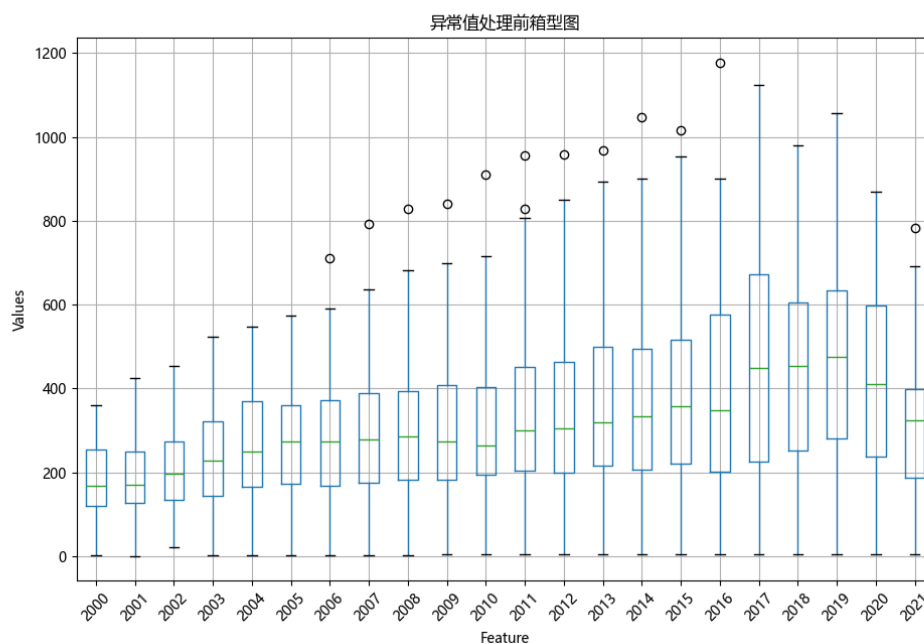


图 4 年份处理后的碳排放箱型图

由图片可以看到图中的各年份的省份碳排放中位数, 在 2017 年前呈上升趋势

势，而到 2017 到 2021 年间中位数呈下降趋势，产生这个现象的原因是我国积极推进减少碳排放的政策。此外，图中离群点的个数较少，可见我国各省份的碳排放量较接近，使得据数据分析得到的数据具有普适性，便于政策的推进。图中所示的碳排放数据的上限与下限之差在逐渐减少。本文在此基础上，抽取了碳排放总量前五的省份进行年份的趋势分析得到下图：



图 5 碳排放量前五的省份年度数据

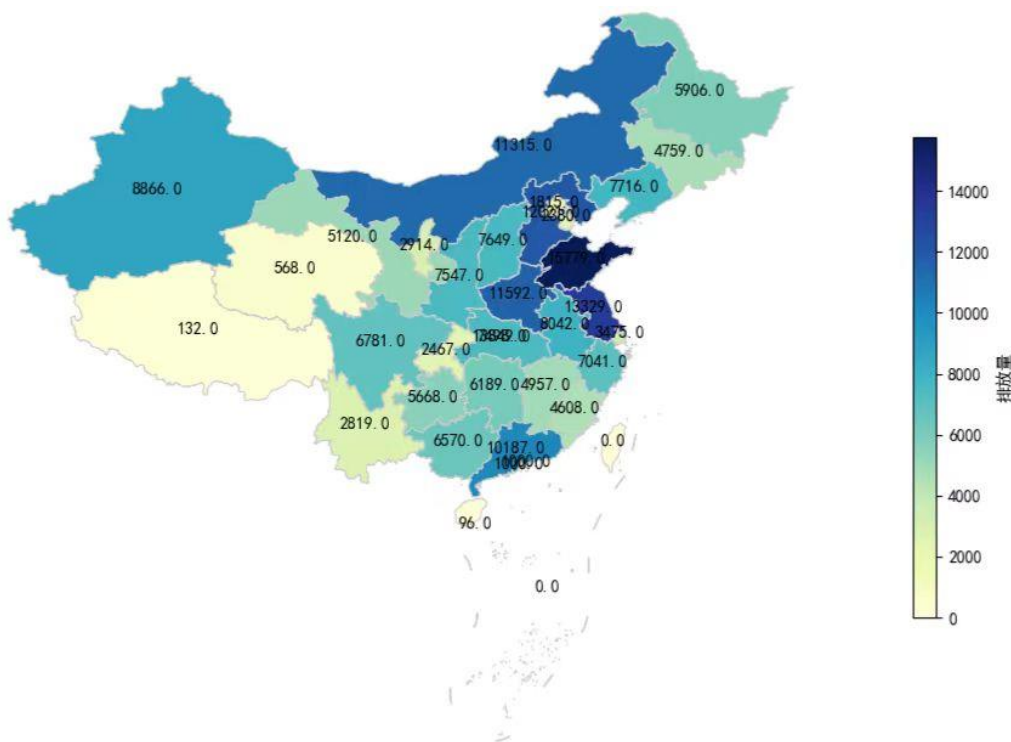


图 6 各省碳排放数据

由图片分析可知各省碳排放量在 2017 至 2021 年间已经呈下降趋势。

3. 2. 2 碳排放数据中各气体的含量数据分析：

在对数据进行初步处理后得到了如下图所示箱型图：

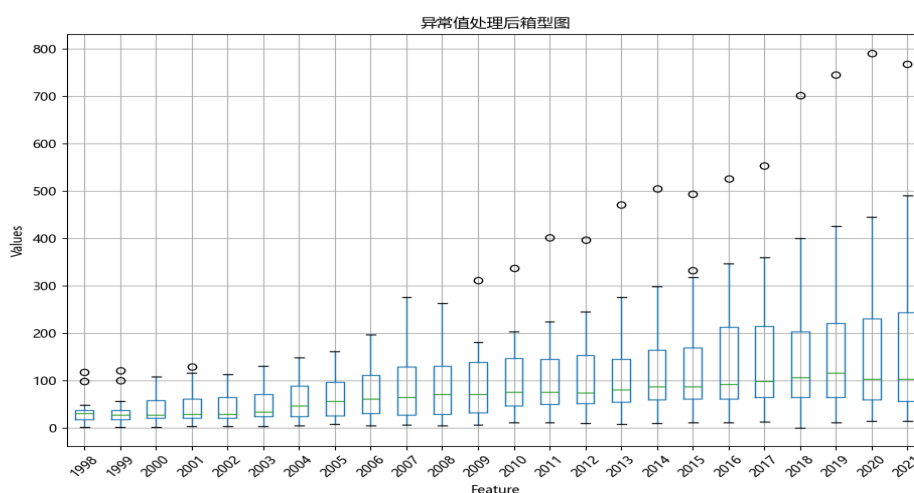


图 7 不同种碳气体的年份箱型图

由图片分析可得，各物质排放量的中位数在 1998 到 2021 年间变化并不大，可见各物质排放量较为稳定，但是由图片分析可知在 1998 到 2021 年间碳化合物排放量的上限与下限之差在逐渐增大，这说明我们需要对化合物间进行关联度分

析和主成分分析。同时在研究时本文对化合物的占比进行了统计结果如下图所示：

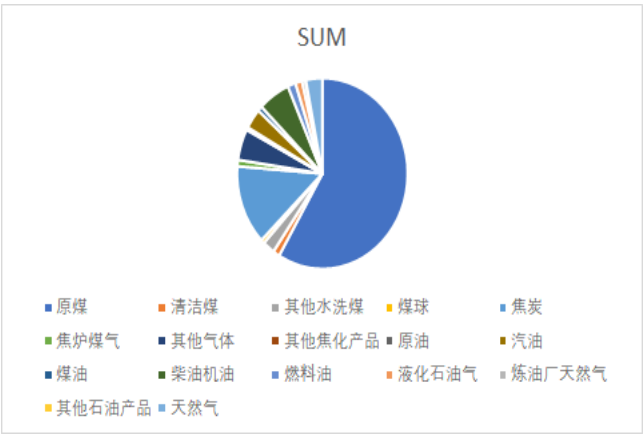


图 8 化合物占比饼状图

据此可得到原煤与焦炭的占比较大，我们需要对其制定方案。并且通过观察我们发现物质种类较多，需要进行主成分分析以减少分析与处理的复杂性。

3. 2. 3 社会经济指标指标数据分析

为了更好的揭示影响碳排放量的因素，我们首先拟合了 2004-2021 间碳排放量和国内生产总值的函数曲线，得到如下结果：

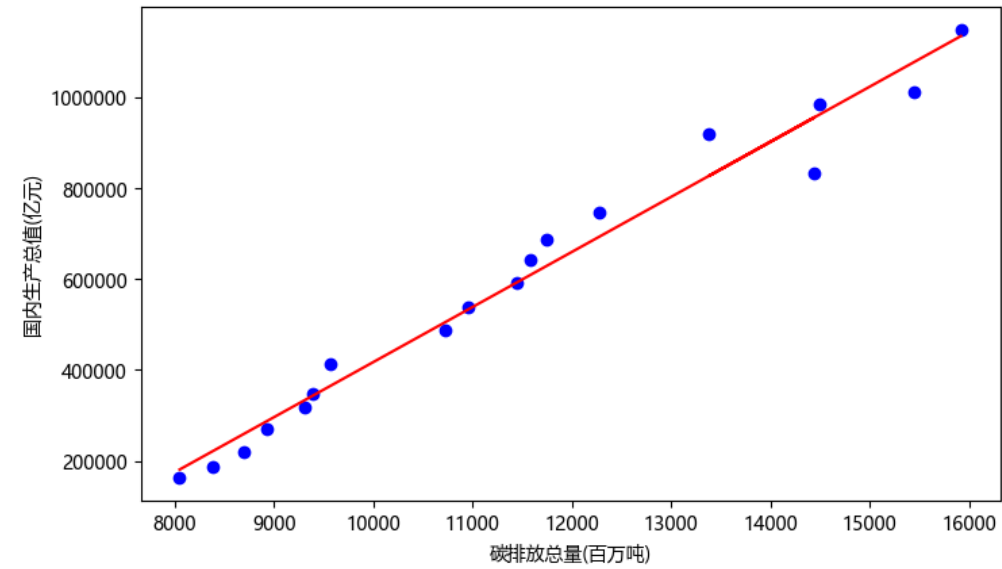


图 9GDP 与碳排放数据的拟合曲线

由图中可以知道 GDP 与碳排放数据成一定的正相关性。为了更好的揭示辅助变量与碳排放量的关系，我们引入了灰色关联度分析法进行了分析。关联度为

两个系统间的因素随时间或不同对象而变化的关联性大小的量度。灰色关联度分析法是衡量因素间的关联程度的一种方法。本文对碳排放数据与辅助变量进行了灰色关联度分析。得到了如下的热力图：

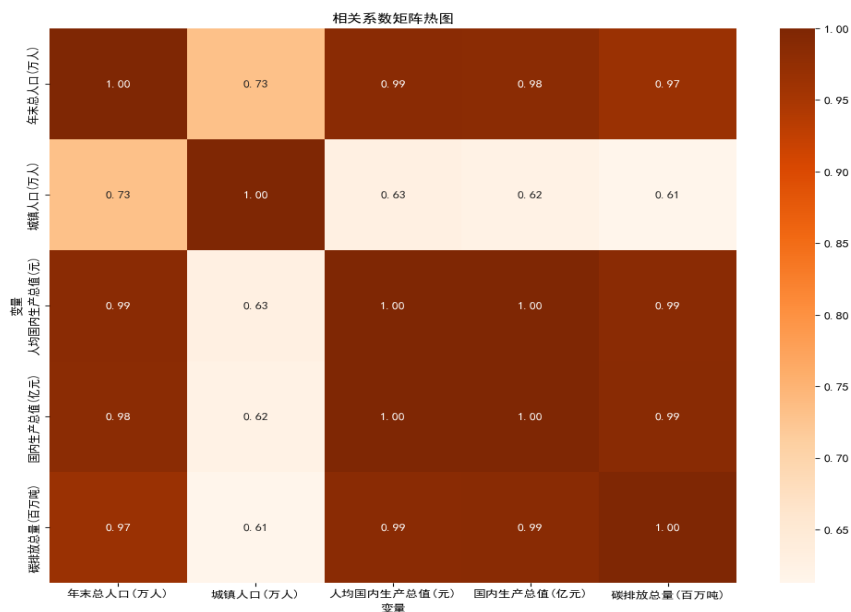


图 10 辅助变量热力图

由图片可以看出碳排放量与国内生产总值的关联性最高为 0.99，和年末人口数，人均国内生产总值也有较强的关联性，与城镇人口的关联性最低为 0.61，目前我国正处于经济发展的重要时期，国内的生产总值与年末人口数仍在增加，这说明我国减排事业仍存在着巨大的挑战。

四. 模型的建立与求解

4.1 碳排放物间的关联分析与主成分分析

4.1.1 碳排放物的主成分分析

4.1.1.1 主成分分析模型介绍

主成分分析（PCA）是一种高效的统计工具，它通过正交变换将一组存在相互关联的变量转换为一组线性无关的变量，即主成分。PCA 作为一项广泛使用的降维技术，主要目的是将高维数据集转换为低维表示，同时尽可能保留原始数据集中的关键信息。该技术的核心目标是识别出一组新的坐标轴，这些坐标轴即为

主成分，它们代表了原始数据中方差最大的方向。通过这种方式，PCA 能够捕捉数据的内在结构，揭示数据中最重要的特征。PCA 算法的一般步骤过程总结见下。

4.1.1.2 标准化

考虑到不同的特征具有不同的单位与尺度，为了防止具有较大尺度的特征对结果产生影响，同时为了提升收敛速度和解释性，应该先对特征量先进行标准化处理。为了进行标准化首先按照列进行计算均值和标准差，计算公式如下其中 S_j 表示标准差：

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (6)$$

$$S_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}} \quad (7)$$

将每个样本的每个特征进行标准化处理，得到标准化特性矩阵

$$X_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (8)$$

标准矩阵如下其中 x_{ij} 表示特征量的值：

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (9)$$

4.1.1.3 计算协方差矩阵

协方差矩阵反映了两两特征之间的线性强度和方向。协方差矩阵提供了特征之间的协方差值，从而反映特征之间的相关性，而主成分分析法的目标正是确定数据中主要的变化方向，这与之不谋而合。同时在进行主成分分析法后得到的特征值具有一定的正交性，而协方差得到的特征具有正交性。因此需要计算协方差矩阵。

协方差矩阵的数学表达如下其中 r_{ij} 表示为协方差：

$$\begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & r_{np} \end{bmatrix} \quad (10)$$

协方差计算公式如下：

$$r_{ij} = \frac{1}{n-1} \sum_{k=1}^n X_{ki} X_{kj} \quad (11)$$

4.1.1.4 计算协方差矩阵的特征值与特征向量

特征值反映了协方差矩阵的变化方向，特征向量是这些变换方向上的向量。据特征值的大小我们可以选择主成分，进而通过主成分进行降维。

4.1.1.5 适合性检验

KMO 检验统计量是用于比较相关系数与偏相关系数的指标，常用于多元统计的因果分析。如果得到的值越接近 1，则说明主成分分析的适用性越强。我们对得到的协方差矩阵进行适合性检验，得到的结果如下：

KMO 和巴特利特检验			
→	KMO 取样适切性量数。		0.722
	巴特利特球形度检验	近似卡方	1453.238
		自由度	190
		显著性	0.000

图 11KMO 检验结果

由检验结果可以看到 kmo 系数为 0.722。说明具有较强的适应性。

4.1.1.6 计算主成分贡献率及累计贡献率

第 i 个主成分的贡献率计算式如下：

$$\frac{\lambda_i}{\sum_{k=1}^p \lambda_k} \quad (12)$$

累计贡献率计算式如下：

$$\frac{\sum_{j=1}^i \lambda_j}{\sum_{k=1}^p \lambda_k} \quad (13)$$

计算结果如下：

```
D:\Desktop\数学建模国赛\Python\Scripts\python.exe D:\Col2_code\Python\main.py
[0.995912 0.002929 0.000628 0.000235 0.00012 0.000083 0.000033 0.000023
0.000013 0.000008 0.000006 0.000004 0.000003 0.000001 0.000001 0.000001
0.      ]
```


图 12 主成分分析结果一

由结果可以看到只有一个结果较大，其余结果较小，这可能是由于该特征的尺度较大对结果产生了影响，为了进一步分析主成分，所以我们去掉了该因素再次做了主成分分析，结果如下：

```
D:\Desktop\数学建模国赛\Python\Scripts\python.exe D:\Col2_code\Python\main.py
[0.849351 0.066094 0.04382 0.018829 0.00787 0.005092 0.003752 0.002025
 0.001161 0.000718 0.000503 0.000282 0.000186 0.000138 0.000095 0.000065
 0.000018]
进程已结束，退出代码为 0
```

图 13 主成分分析结果二

由主成分分析后得到的结果可以看出原煤，焦炭和柴油机油为主成分，我们需要重点监测这三种物质的排放情况并且需要制定切实可行的策略。

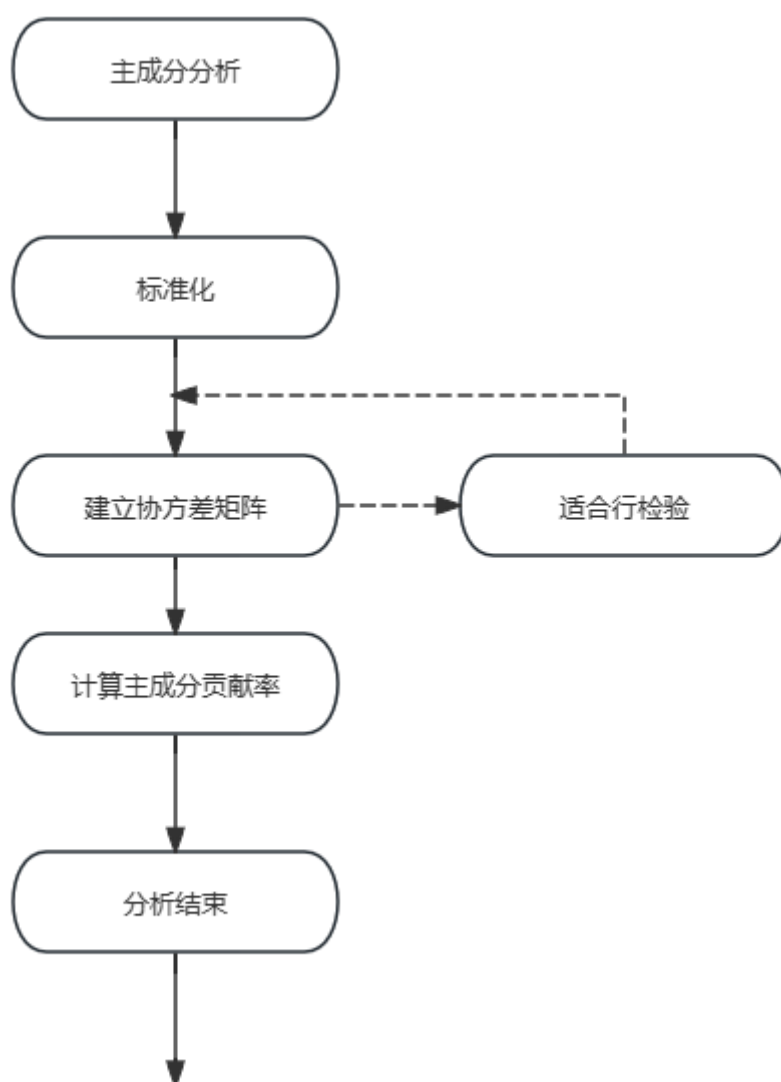


图 14 主成分分析流程图

4.1.2 碳排放物间的关联分析

4.1.2.1 灰色关联度分析介绍

灰色关联度分析是一项多因素统计分析的方法，被广泛用于评估灰色系统中各指标与因素之间的相关性。其独特之处在于，通过比较参考数据列和若干比较数据列的几何形状相似程度，来判断它们之间的联系是否紧密。这种方法能够有效地反映出曲线间的关联程度，帮助我们更清晰地理解各指标之间的关系。灰色关联度分析的应用不仅局限于特定领域，而且在不同行业和学科中都具有广泛的适用性。通过深入研究和运用这一方法，我们可以更准确地评估和预测各种复杂系统中的变量之间的相互作用，从而为决策制定和问题解决提供更可靠的依据。

4.1.2.2 灰色关联度结果分析

我们通过 `python` 计算了各变量间的灰色相关度，并且绘制了相关系数热力图，得到的结果如下：

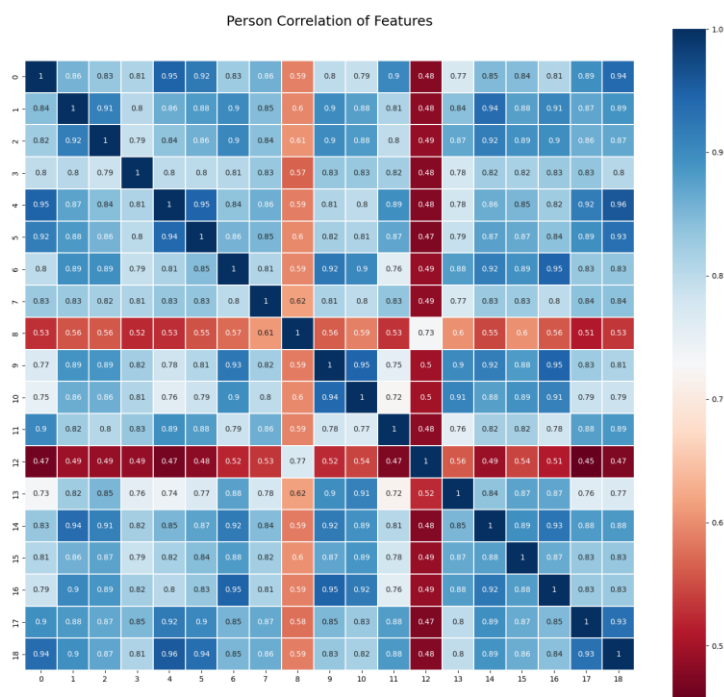


图 15 因素间的灰色关联度分析

由热力图可知,天然气与原煤的关联性较强,焦炭与柴油机油的关联性较强,焦炭与清洁煤的关联性较强。在制定政策时可以重点的考虑相关性较强的物质。

4.2 碳排放预测模型的建立

4.2.1 模型介绍

ARIMA 模型为差分整合移动平均自回归模型又称整合移动自规模型,是一种进行时间序列预测的常用的分析方法。ARIMA 模型结合了自回归 (AR), 差分 (I), 移动平均 (MA) 这三个部件。自回归描述了时间序列与过去值的关系, 差分的作用使时间序列数据平稳, 移动平均描述了时间序列数据与过去误差项间的关系。因此 ARIMA 模型也常表示为 ARIMA(p,d,q), 其中 p 为自回归的阶数, d 为差分阶数, q 为移动平均的阶数, 即在模型拟合时考虑前 q 个误差的影响。如何确定这三个值极为模型拟合的关键。在本文中我们将其引入来对未来年份的碳排放量进行预测。

4.2.2 平稳性检验

ADF 检验是一种常用的时间序列分析工具,用于检验时间序列数据是否具有单位根,是评估时间序列数据平稳性的重要工具,可以帮我们更好的理解和分析时间序列数据的特性与行为。所以在进行时间序列预测时首先我们需要利用 ADF 检验法进行检验,而在 ADF 检验时,本文引入了常用的两个关键统计量 p 值与检验统计量(value)值, p 值是一个介于 0 与 1 的数值,用来判断检验的统计显著性。value 值是 ADF 检验的统计量,与 p 值一起使用来做出关于时间序列平稳性的决策。在进行检验后如果得到的 p 值与 value 值之差大于 0.05,我们需要进行差分运算,直至 p 值与 value 值之差小于 0.05,差分运算是时间序列分析中常用的一种方法,用于消除时间序列数据中的趋势成分,从而使得时间序列更加平稳,使得得到的时间序列模型在进行预测时更加准确。差分运算的数学表达式如下,其中 x_t 表示原始时间序列, y_t 表示变换后的时间序列:

$$y_t = x_t - x_{t-1} (t = 2, 3, \dots, 22) \quad (4)$$

一次差分运算计算式, 其中 x_t 为时间序列。

$$y_2^2 = y_t - y_{t-1} = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2} \quad (5)$$

二次差分运算计算式。

$$y_{n_t} = y(n-1)_t - y(n-2)_t \quad (6)$$

n 次差分运算计算式。

利用 python 进行 ADF 检验得到的结果如下：

```
-----
ADF检验结果：
ADF Statistic: 0.387347
p-value: 0.981016
Lags Used: 5
Observations Used: 14
Critical Values:
    1%: -4.012
    5%: -3.104
   10%: -2.691
-----

差分后序列的ADF检验结果：
ADF Statistic: -6.748537
p-value: 0.000000
Lags Used: 0
Observations Used: 18
Critical Values:
    1%: -3.859
    5%: -3.042
   10%: -2.661
-----
```

图 16 ADF 检验结果

由计算结果可以看到在未进行差分运算时 p-value 的值大于 0.05，而进行一次差分后得到的 p-value 值小于 0.05。因此本文选择了一次差分。同时在研究时我们构建了动态化差分调整，这可以在差分运算时根据拟合效果动态的调整差分次数，可以进行更加准确的预测。

4.2.3 模型拟合确定最优 p, q 值

表 1 样本自相关系数与偏相关系数

	AR(p)	MA(q)	ARMA(p,q)
ACF	拖尾	滞后 q 阶后截尾	拖尾
PACF	滞后 p 阶后截尾	拖尾	拖尾

样本自相关系数（ACF）和偏自相关系数（PACF）在识别纯自回归（AR）或移动平均（MA）模型方面发挥着关键作用，但对于混合自回归移动平均（ARMA）模型，由于其复杂性，可能会出现混合信号，使得模型识别变得更加棘手。为了克服这一挑战，我们转向 AIC（赤池信息量准则）和 BIC（贝叶斯信息准则），这两者是常用的模型选择准则。AIC 的目标是在寻找最佳拟合数据的同时避免模型过度拟合，它平衡了模型的拟合优度和模型的复杂度，提供了一个量化的方法来评估不同模型之间的相对优劣。它的计算公式如下：

$$2k - 2\ln(L) \quad (7)$$

其中 k 是模型中参数的数量， L 为模型的似合函数。

BIC 与 AIC 较为相似，但在选择模型时存在不同，BIC 会更加重视样本数量，对模型复杂度的惩罚更严格。他的计算公式如下：

$$\ln(n)k - 2\ln(L) \quad (8)$$

其中 n 为样本容量。选取 p 与 q 值是较为关键的步骤，我们使用了 AIC 和 BIC 法确定了较为合适的 p 与 q 值，结果如下图：

```
the AIC is(0, 1)
the BIC is(0, 1)
```

图 17 BIC 与 AIC 检验结果

通过 BIC 与 AIC 确定最优的 p, q 值分别为 0 和 1，因此确定了 ARIMA 模型的预取值为 1，0，1，模型的运行结果如下图所示：

```

-----
ma.L1          -0.999218
sigma2         408422.661204
dtype: float64

SARIMAX Results
=====
Dep. Variable:          全国碳排放量    No. Observations:          19
Model:                ARIMA(0, 1, 1)    Log Likelihood              -143.145
Date:                 Fri, 12 Apr 2024    AIC                        290.289
Time:                 21:04:55            BIC                        292.070
Sample:               0                  HQIC                      290.535
                  - 19
Covariance Type:      opg
=====

```

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.9992	0.242	-4.137	0.000	-1.473	-0.526
sigma2	4.084e+05	5.91e-07	6.91e+11	0.000	4.08e+05	4.08e+05

```

=====
Ljung-Box (L1) (Q):          4.48    Jarque-Bera (JB):          2.64
Prob(Q):                    0.03    Prob(JB):                0.27
Heteroskedasticity (H):      8.41    Skew:                    0.25
Prob(H) (two-sided):         0.02    Kurtosis:                4.81
=====

```

图 18 模型运行结果

4.2.4 模型的检验

在回归分析 $y = \beta_0 + \beta_1 x + \varepsilon$ 中，假定 ε 的期望值为 0，方差相等且服从正态分布的一个随机变量。但是，若关于 ε 的假定不成立，此时所做的检验以及估计和预测也许站不住脚。确定有关的 ε 假定是否成立的方法之一是进行残差分析。我们对建立的模型进行残差检验得到的双 Q 图如下所示：

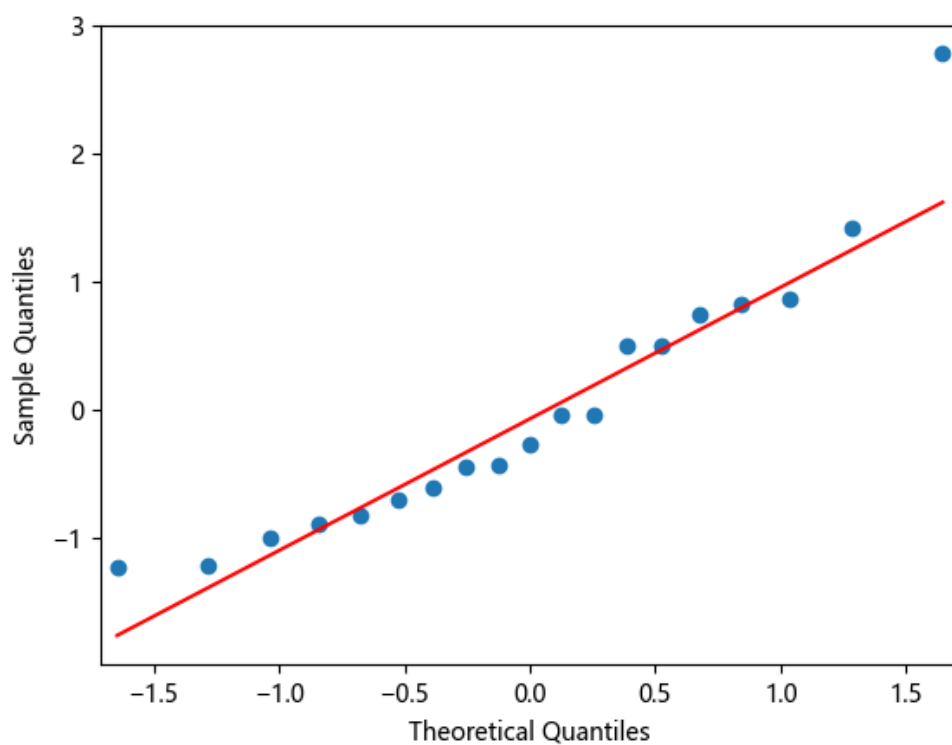


图 19 残差检验双 Q 图

由图中信息所示，散点均匀的分布在直线附近说明残差序列接近正太分列，相互独立。证明了模型的拟合效果较好。

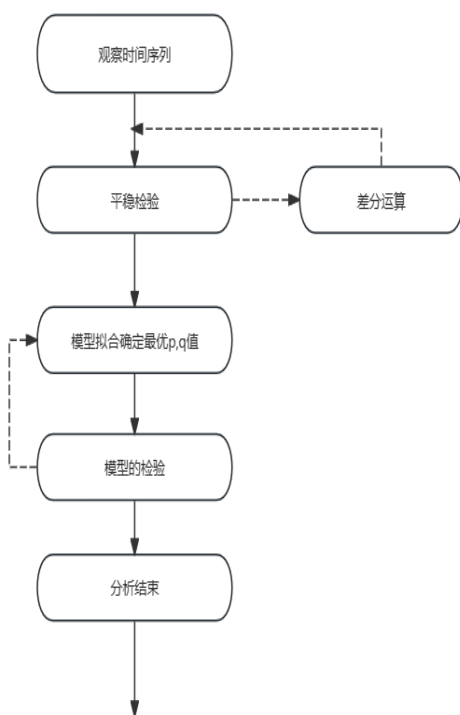


图 20 ARIMA 流程图

4.2.5 碳排放量预测和结果分析

在之前我们已经建立了 `arima` 时间预测模型，通过编程我们得到了 2020 至 2021 的碳排放量预测值。我们将其与所获得的值进行对比，得到下图：

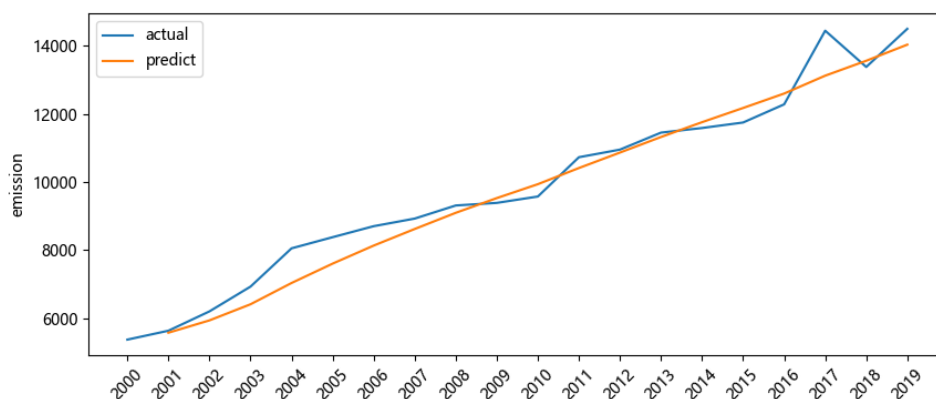


图 21 实际与预测对比图

由图片所示，在 2000 至 2019 年间，碳排放的预计值和实际值相差并不大，且具有相同的趋势。也再次证明了预测值的准确性。通过图中观察在 2017 年至 2018 年间，碳排放量已经有了明显降低，但 2018 至 2019 年间碳排放量有回升趋势，这提醒我们需要注意减少碳排放量的措施和决策的制定。

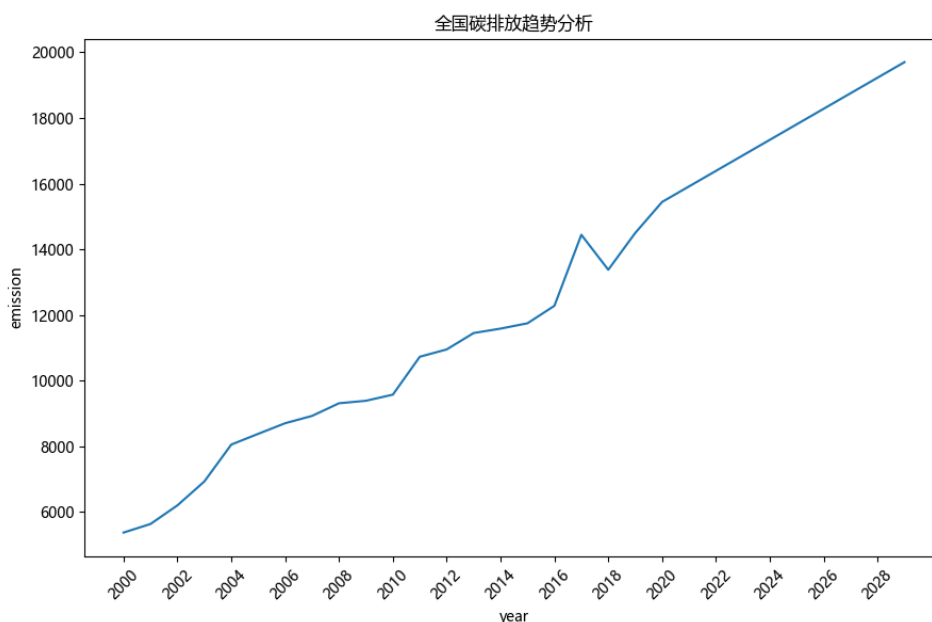


图 22 碳排放量的整合

全国碳排放量		mean	mean_se	mean_ci_lower	mean_ci_upper
19	14970.439951	15152.524542	13686.310156	16254.569746	
20	15443.535663	15625.620444	14159.405494	16727.665832	
21	15916.631374	16098.716346	14632.500831	17200.761917	
22	16389.727086	16571.812249	15105.596169	17673.858003	
23	16862.822797	17044.908151	15578.691506	18146.954088	
24	17335.918509	17518.004053	16051.786844	18620.050174	
25	17809.014220	17991.099955	16524.882181	19093.146259	
26	18282.109931	18464.195858	16997.977519	19566.242344	
27	18755.205643	18937.291760	17471.072856	20039.338430	
28	19228.301354	19410.387662	17944.168193	20512.434515	
29	19701.397066	19883.483564	18417.263531	20985.530601	

图 23 碳排放的预测值

碳排放物主成分的排放量预测图如下：

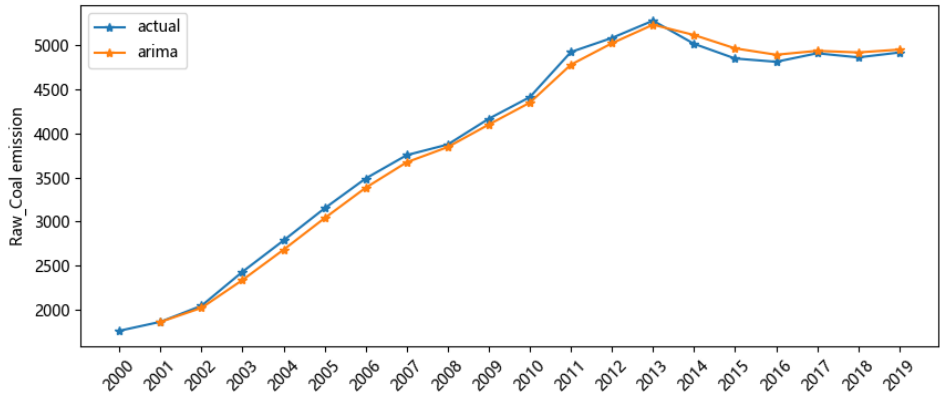


图 24 原煤排放量预测图

Raw_Coal	mean	mean_se	mean_ci_lower	mean_ci_upper
19	4944.315260	5083.668107	4607.557959	5281.072561
20	4976.780520	5153.291352	4567.194908	5386.366133
21	5009.245780	5217.239564	4537.954715	5480.536846
22	5041.711041	5277.518210	4515.906740	5567.515341
23	5074.176301	5335.172523	4499.002365	5649.350237
24	5106.641561	5390.829402	4486.012887	5727.270235
25	5139.106821	5444.899941	4476.132579	5802.081063
26	5171.572081	5497.671008	4468.799188	5874.344974
27	5204.037341	5549.352204	4463.601906	5944.472776
28	5236.502602	5600.102192	4460.229758	6012.775445
29	5268.967862	5650.044511	4458.440612	6079.495111

图 25 原煤排放量预测

通过上述的两个图片我们可以明显地看到，原煤的排放量从 2000 年到 2013

年间呈现了较强的增长趋势，而之后虽然呈现增长趋势，但有了增长的数度有了明显的下降趋势。

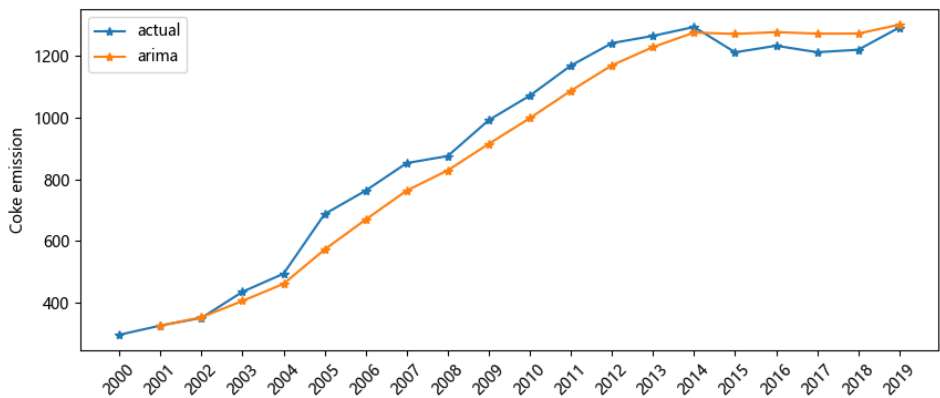


图 26 焦炭排放量预测图

Diesel_Oil	mean	mean_se	mean_ci_lower	mean_ci_upper
19	429.201758	477.275844	392.264582	466.138934
20	399.973515	450.649158	357.937383	442.009648
21	370.745273	423.739376	324.165042	417.325503
22	341.517030	396.622700	290.798204	392.235857
23	312.288788	369.346267	257.744487	366.833089
24	283.060546	341.941638	224.942026	341.179065
25	253.832303	314.431180	192.346990	315.317616
26	224.604061	286.831422	159.926977	289.281145
27	195.375818	259.154994	127.657234	263.094402
28	166.147576	231.411802	95.518346	236.776806
29	136.919333	203.609787	63.494749	210.343917

图 27 焦炭排放量预测

通过图片显示在 200 年~2013 年间呈现增长率逐渐增加的上升趋势，在 2014 年间有下降趋势但 2015 年之后呈现了一定的上升趋势

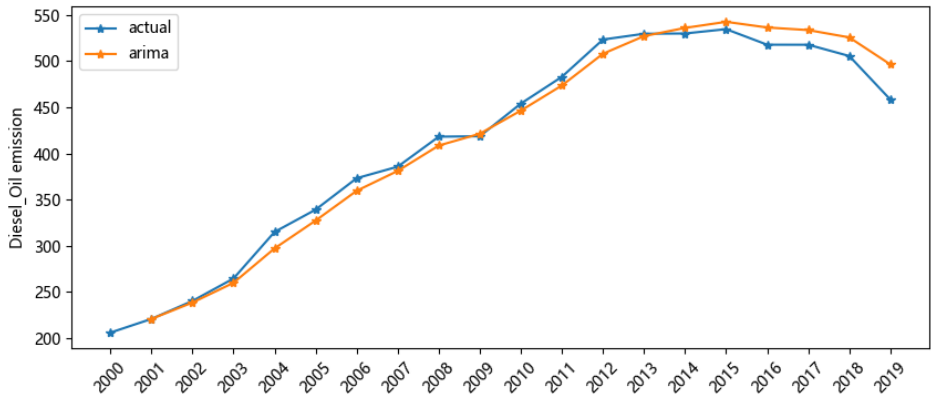


图 28 柴油机油排放量预测图

Diesel_Oil	mean	mean_se	mean_ci_lower	mean_ci_upper
19	429.201758	477.275844	392.264582	466.138934
20	399.973515	450.649158	357.937383	442.009648
21	370.745273	423.739376	324.165042	417.325503
22	341.517030	396.622700	290.798204	392.235857
23	312.288788	369.346267	257.744487	366.833089
24	283.060546	341.941638	224.942026	341.179065
25	253.832303	314.431180	192.346990	315.317616
26	224.604061	286.831422	159.926977	289.281145
27	195.375818	259.154994	127.657234	263.094402
28	166.147576	231.411802	95.518346	236.776806
29	136.919333	203.609787	63.494749	210.343917

图 29 柴油机油排放量预测

通过图片可以看到在 2000 年~2012 年阶段柴油机油的增长率逐渐增加且有上升趋势但在此之后呈现下降趋势

五. 结论

5.1 结论一

在本文中为了预测未来的碳排放量我们建立了 ARIMA 时间序列预测模型，结合了中国碳核算数据库的《1997-2019 年 290 个中国城市碳排放数据》进行了碳排放量的预测，经过研究我们发现碳排放量仍呈上升趋势，且上升速度具有较快的增涨速度，如果想要在 2030 年达到碳排放量峰值，我们仍需要制定切实可行的政策，并且需要密切关注碳的排放量，同时在本文中我们通过主成分分析法选出了主成分并且通过时间序列预测法进行了主成分的排放量预测。

5.2 结论二

为了探究碳排放物之间的相关性和数量众多的碳排放物中的主成分。我们建立了主成分分析和灰色关联度分析模型。在进行主成分分析后我们发现原煤，焦炭和柴油机油这三中碳排放物为主成分，这告诉我们在关注碳排放量时可以重点监测这三种物质的总量与变化趋势。同时在本文中提供了主要碳排放物间的相关系数热力图，这可以帮助我们制定政策更加精确有效，使政策具有更多的针对性，

如本文中分析得天然气与原煤的关联性较强，焦炭与柴油机油的关联性较强，焦炭与清洁煤的关联性较强。同时得到的主成分可以通过时间序列预测法进行预测分析。

六. 建议与启示

尽管近年来中国在减少碳排放方面取得了一定的进展，但仍面临着诸多挑战。目前，我国的碳排放量依然较高，且主要依赖于煤炭等传统能源，这不仅对环境造成了压力，也限制了我国经济的可持续发展。因此，为了实现 2030 年前碳排放达到峰值的目标，并进一步推动可持续发展，我们需要采取更为全面和深入的措施。结合以上实证分析和研究结果，我们提出以下建议：

6.1 绿色经济发展策略

鉴于 GDP 与碳排放量之间的正相关性，建议在经济发展策略中更加注重绿色增长和可持续发展。这可以通过推广循环经济、绿色制造和节能减排等措施来实现。政府应当引导资金流向低碳环保项目，支持绿色金融的发展，同时鼓励企业通过创新实现经济增长与环境保护的双赢。此外，加强公众教育，提高社会对可持续发展的认识和参与度，共同推动经济社会向绿色转型。

6.2 能源结构优化策略

对于目前主要碳排放物之间的关系，我们可以通过制定和执行更为严格的排放标准，激励企业积极探索和采用清洁能源，逐步减少对原煤的依赖。同时，我们应当着力提高天然气的使用效率，并在工业生产与日常生活中推广其应用，以此作为减少碳排放的有效途径。

6.3 工业生产领域技术革新策略

对工业生产领域进行技术革新也是至关重要的一环。通过推动工业炉的改造升级，例如采用电弧炉等环保技术，可以有效降低焦炭等传统燃料的使用，从而

减少碳排放。同时，加强对柴油机油的质量和使用的监管，推广清洁燃料和高效能效的交通工具，也是降低交通运输领域碳排放的有效手段。

6.4 燃料使用优化战略

对于煤炭行业，我们可以鼓励并支持煤矿企业进行技术升级，提高清洁煤的生产和使用比例。同时，加大对清洁煤技术研发的投入，不断提高清洁煤的燃烧效率，以实现更高效的能源利用和更少的碳排放。

我们认识到，由于实际情况的复杂性和多变性，这些建议可能需要进一步的调整和优化，以确保其有效性和适应性。

七. 附录

附录 1
ARIMA 预测碳排放以及模型预测
<pre># Step 1: 数据预处理 import pandas as pd import statsmodels.tsa.stattools as ts #adf 检验 from matplotlib import pyplot as plt from sklearn.metrics import mean_squared_error, mean_absolute_error, median_absolute_error from statsmodels.graphics.tsaplots import plot_acf,plot_pacf #ACF 与 PACF from statsmodels.tsa.arima.model import ARIMA #ARIMA 模型 import statsmodels.api as sm from statsmodels.graphics.api import qqplot #qq 图 import numpy as np from sklearn.svm import SVR from sklearn.model_selection import train_test_split df=pd.read_csv("C:/Users/Administrator/Desktop/无标题 4.csv",encoding='gbk') print(df["全国碳排放量"]) # 创建一个 DatetimeIndex 对象, 指定时间范围为 2021 年 1 月 1 日到 1 月 5 日 dates = pd.date_range(start='2000-01-01', end='2019-12-31', freq='AS') print(dates) # # 创建一个 Series 对象, 将列表作为数据, DatetimeIndex 对象作为索引 t_s = pd.Series(data=df["全国碳排放量"].tolist(), index=dates)</pre>

```

t_s.index = t_s.index.strftime('%Y')
# print(t_s.index)
plt.rcParams['font.sans-serif'] = ['Microsoft YaHei']
plt.figure(figsize=(13,7))
plt.plot(t_s.index, t_s)
tick_positions = t_s.index[::2]
plt.xticks(tick_positions,rotation=45)
plt.xlabel("year")
plt.ylabel("emission")
plt.title("全国碳排放趋势分析")
plt.show()

# 进行ADF 检验
result = ts.adfuller(df['全国碳排放量'])
# 输出 ADF 结果
print('-----')
print('ADF 检验结果: ')
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Lags Used: %d' % result[2])
print('Observations Used: %d' % result[3])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))

##### 差分运算
def diff(timeseries):
    new_df=timeseries.diff(periods=1).dropna() #dropna 删除 NaN
    # sec_df=new_df.diff(periods=1).dropna()
    new_df.plot(color='orange',title='差分运算')
    plt.show()
    return new_df

#进行一阶差分
ndf=diff(df['全国碳排放量'])

#再次进行 ADF 检验
result2 = ts.adfuller(ndf)
# 输出 ADF 结果
print('-----')
print('差分后序列的 ADF 检验结果:')
print('ADF Statistic: %f' % result2[0])
print('p-value: %f' % result2[1])
print('Lags Used: %d' % result2[2])

```

```

print('Observations Used: %d' % result2[3])
print('Critical Values:')
for key, value in result2[4].items():
    print('\t%s: %.3f' % (key, value))

#模型选择: 绘制 ACF 与 PACF, 即自相关图和偏自相关图
#### 绘制 ACF 与 PACF 的图像
def plot_acf_pacf(timeseries): #利用 ACF 和 PACF 判断模型阶数
    plot_acf(timeseries, lags=timeseries.shape[0]%2) #延迟数
    plot_pacf(timeseries, lags=timeseries.shape[0]%2)
    plt.show()
plot_acf_pacf(ndf)

#迭代调优
print('-----')
#AIC
timeseries=ndf
AIC=sm.tsa.stattools.arma_order_select_ic

    (timeseries, max_ar=4, max_ma=4, ic='aic')['aic_min_order']
#BIC
BIC=sm.tsa.stattools.arma_order_select_ic

    (timeseries, max_ar=4, max_ma=4, ic='bic')['bic_min_order']
print('the AIC is{}\nthe BIC is{}\n'.format(AIC,BIC))

#模型构建
print('-----')
model= ARIMA(ndf, order=(0,1,1)).fit()
print(model.params)
print(model.summary())

#残差图
model.resid.plot(figsize=(10,3))
plt.title("残差图")
plt.show()

#QQ 图看正态性
qqplot(model.resid, line="q", fit=True)
plt.show()

#模型预测
print('-----')
predict= model.predict(1,19) #dynamic=True)
# print('模型预测:', '\n', predict)

```

```

#反向差分运算
# 对差分后的时间序列进行逆差分运算, 两个参数: 差分数据序列和原始数据序列
def inverse_diff(diff_series, original_series):
    inverted = []
    prev = original_series.iloc[0]
    for val in diff_series:
        current = val + prev
        inverted.append(current)
        prev = current
    return pd.Series(inverted, index=original_series.index[1:])
n_predict=inverse_diff(predict,t_s)
print('模型预测:', '\n', n_predict)

# #画图
plt.figure(figsize=(10,4))
plt.plot(t_s.index,t_s,label='actual', linestyle='-',marker='*')
plt.plot(predict.index,n_predict,label='predict')

plt.xticks(t_s.index,rotation=45)
plt.legend(['actual','arima'])
plt.xlabel('time(year)')
plt.ylabel('emission')
plt.show()

print('-----')
# 进行三步预测, 并输出 95%置信区间
steps=11 #未来三期预测
forecast= model.get_forecast(steps=steps)
table=pd.DataFrame(forecast.summary_frame())
# print(table.iloc[1])
table.iloc[0]=table.iloc[0]+t_s[-1]
# print(table.iloc[0, 0])
for i in range(steps-1):
    table.iloc[i+1]=table.iloc[i+1]+table.iloc[i, 0]
print(table)

```

附录 2

全国碳排放分布图绘制


```

import pandas as pd
import geopandas as gpd

import numpy as np
import matplotlib.pyplot as plt

# 指定中文字体路径
plt.rcParams['font.sans-serif'] = ['SimHei'] # 指定中文字体为黑体
plt.rcParams['axes.unicode_minus'] = False # 解决负号显示问题

# 然后绘制图像
# 读取地理数据
gdf = gpd.read_file("C:\\Users\\13508\\Desktop\\中华人民共和国",
encoding='utf-8')
sales_data = pd.read_excel("C:\\Users\\13508\\Desktop\\新建 XLSX 工作表.xlsx")
ind = sales_data['name']
sales_data = sales_data.drop(columns='name')
gdf.to_excel("C:\\Users\\13508\\Desktop\\工作簿 2.xlsx")
df = pd.DataFrame()
df['name'] = ind.values
df['value'] = np.floor(sales_data.sum(axis=1).values)

# 合并销量数据到地理数据中
merged_data = gdf.merge(df, how='left', left_on='name', right_on='name')

# 绘制地图
fig, ax = plt.subplots(figsize=(12, 10))
ax.set_title('全国排放量分布图', fontproperties='SimHei')

# 更改地图背景颜色
ax.set_facecolor('lightblue')

merged_data.plot(column='value', cmap='YlGnBu', linewidth=0.8, ax=ax,
edgecolor='0.8', legend=True,
legend_kwds={'label': '排放量', 'orientation': "vertical",
'shrink': 0.6})

# 在地图上添加省份碳排放量标签
for idx, row in merged_data.iterrows():
    plt.annotate(text=row['value'], xy=(row['geometry'].centroid.x,
row['geometry'].centroid.y),
horizontalalignment='center', fontproperties='SimHei')

ax.set_axis_off()

```

```
plt.show()
```

附录 3

异常值处理及箱型图

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

# 箱型图判断异常点
def box_outlier(data):
    df = data.copy(deep=True)
    out_index = []
    for col in df.columns: # 对每一列分别用箱型图进行判断
        Q1 = df[col].quantile(q=0.25) # 下四分位
        Q3 = df[col].quantile(q=0.75) # 上四分位
        low_whisker = Q1 - 1.5 * (Q3 - Q1) # 下边缘
        up_whisker = Q3 + 1.5 * (Q3 - Q1) # 上边缘

        # 寻找异常点,获得异常点索引值, 删除索引值所在行数据
        rule = (df[col] > up_whisker) | (df[col] < low_whisker)
        out = df[col].index[rule]
        out_index += out.tolist()

    df.drop(out_index, inplace=True)
    return df

def getFc(data, first_column):
    df = data.copy(deep=True)
    fc = first_column.copy(deep=True)
    out_index = []
    for col in df.columns: # 对每一列分别用箱型图进行判断
        Q1 = df[col].quantile(q=0.25) # 下四分位
        Q3 = df[col].quantile(q=0.75) # 上四分位
        low_whisker = Q1 - 1.5 * (Q3 - Q1) # 下边缘
        up_whisker = Q3 + 1.5 * (Q3 - Q1) # 上边缘

        # 寻找异常点,获得异常点索引值, 删除索引值所在行数据
        rule = (df[col] > up_whisker) | (df[col] < low_whisker)
        out = df[col].index[rule]
        out_index += out.tolist()

    df.drop(out_index, inplace=True)
```

```

    fc.drop(out_index, inplace=True)
    return fc

# 读取数据
data = pd.read_csv("C:/Users/13508/Desktop/无标题.csv").dropna(axis=0) # 确保
删除 NaN 是适当的
numeric_cols = data.iloc[:, 1:].select_dtypes(include=[np.number]).columns
numeric_data = data[numeric_cols]
first_column = data.iloc[:, 0]
# 去除异常值
cleaned_data = box_outlier(numeric_data)
new_first_column = getFc(numeric_data, first_column)
new_data = pd.concat([new_first_column, cleaned_data], axis=1)
new_data.to_csv("C:/Users/13508/Desktop/cleaned_data.csv",
index=False, encoding='gbk')
# 绘制箱线图
plt.figure(figsize=(11, 7))
numeric_data.boxplot()
plt.rcParams['font.sans-serif'] = ['Microsoft YaHei']
plt.title('异常值处理前箱型图')
plt.xlabel('Feature')
plt.ylabel('Values')
plt.xticks(rotation=45)
plt.show()

plt.figure(figsize=(11, 7))
cleaned_data.boxplot()
plt.title('异常值处理后箱型图')
plt.xlabel('Feature')
plt.ylabel('Values')
plt.xticks(rotation=45)
plt.show()

```

附录 4

灰色关联度分析

```

#灰色关联度分析
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

wine = pd.read_csv("D:\Col2_code\Python\灰色关联度分析.csv", index_col=0)
# print(wine)

#无量纲化
def dimensionlessProcessing(df_values, df_columns):
    from sklearn.preprocessing import StandardScaler
    scaler = StandardScaler()
    res = scaler.fit_transform(df_values)
    return pd.DataFrame(res, columns=df_columns)

# 求第一列(影响因素)和其它所有列(影响因素)的灰色关联值
def GRA_ONE(data, m=0): # m为参考列
    # 标准化
    data = dimensionlessProcessing(data.values, data.columns)
    # 参考数列
    std = data.iloc[:, m]
    # 比较数列
    ce = data.copy()

    n = ce.shape[0]
    m = ce.shape[1]

    # 与参考数列比较, 相减
    grap = np.zeros([n, m])
    for i in range(m):
        for j in range(n):
            grap[j, i] = abs(ce.iloc[j, i] - std[j])

    # 取出矩阵中的最大值和最小值
    mmax = np.amax(grap)
    mmin = np.amin(grap)
    ρ = 0.5 # 灰色分辨系数

    # 计算值
    grap = pd.DataFrame(grap).apply(lambda x: (mmin + ρ * mmax) / (x + ρ * mmax))

```

```

    # 求均值, 得到灰色关联值
    RT = grap.mean(axis=0)
    return pd.Series(RT)

# 调用 GRA_ONE, 求得所有因素之间的灰色关联值
def GRA(data):
    list_columns = np.arange(data.shape[1])
    df_local = pd.DataFrame(columns=list_columns)
    for i in np.arange(data.shape[1]):
        df_local.iloc[:, i] = GRA_ONE(data, m=i)
    return df_local

data_gra = GRA(wine)
data_gra.to_csv('new_data.csv', index=False)
print(data_gra)

def ShowGRAHeatMap(data):
    # 色彩集
    colormap = plt.cm.RdBu
    plt.figure(figsize=(18, 16))
    plt.title('Person Correlation of Features', y=1.05, size=18)
    sns.heatmap(data.astype(float), linewidths=0.1, vmax=1.0, square=True, \
                cmap=colormap, linecolor='white', annot=True)
    plt.show()

ShowGRAHeatMap(data_gra)

# 行列互换
import pandas as pd
data = pd.read_csv("D:\Col2_code\Python\制作箱型图.csv")
transposed_data = data.transpose()

transposed_data.to_csv("the new transation.csv", index=False)

```

八. 参考文献

- [1] 刘鹏飞, 韩晓琳. 数字技术发展对区域碳排放的影响效应及其作用机制——以长江经济带为例[J]. 生态经济, 2024, 40(04): 26-35.
- [2] 《中国碳达峰碳中和进展报告(2023)》
- [3] 余碧莹, 赵光普, 安润颖, 陈景明, 谭锦潇, 李晓易. 碳中和目标下中国碳排放路径研究[J]. 北京理工大学学报(社会科学版), 2021, 23(2): 17-24. DOI: 10.15918/j.jbitss1009-3370.2021.7380
- [4] 张帆, 徐宁, 吴锋. 共享社会经济路径下中国 2020—2100 年碳排放预测研究. 生态学报, 2021, 41(24): 9691-9704.
- [5] 李永明, 张明. 《碳达峰、碳中和背景下江苏工业面临的挑战、机遇及对策研究》. 现代管理科学. 2021.
- [6] 胡孔玉, & 王竞钊. (无日期). 基于主成分分析法的城市化水平演化研究. 武汉工程大学管理学院, 上海工程技术大学.
- [7] 马晓龙. (2006). 旅游业发展影响因子灰色关联分析. 人文地理, 21(2), 37-40.
- [8] 刘琰. (2023). 基于 ARIMA 模型和指数平滑模型对我国货物进出口总额的预测研究. 统计与决策, 39(2), 45-49.