知网个人查重服务报告单(全文标明引文)

报告编号:BC20230526180446371760873 4　　检测时间:2023-05-26 18:04:46

篇名：基于聚类和惩罚回归的四川省碳排放实证分析

作者：张轩铭;王晓雪;陈勇杭

检测类型：学术出版

比对截止日期：2023-05-26

检测结果

去除本人文献复制比：2.9%　　去除引用文献复制比：2.9%　　总文字复制比：2.9%

单篇最大文字复制比：0.8%（数据驱动的白鹳行为模式挖掘）

重复字符数：[465]　　单篇最大重复字符数：[127]　　总字符数：[16261]

4%(390)　4%(390)　基于聚类和惩罚回归的四川省碳排放实证分析_第1部分（总9813字）

1.2%(75)　1.2%(75)　基于聚类和惩罚回归的四川省碳排放实证分析_第2部分（总6448字）

（注释：无问题部分　文字复制部分　引用部分）

1. 基于聚类和惩罚回归的四川省碳排放实证分析_第1部分　　总字符数：9813

相似文献列表

去除本人文献复制比：4%(390)　　去除引用文献复制比：4%(390)　　文字复制比：4%(390)

1	数据驱动的白鹳行为模式挖掘 曹亚媛(导师：王亮) - 《西安科技大学硕士论文》- 2021-06-01	1.3%（127） 是否引证：否
2	FDI对我国区域碳排放的影响研究 高麟(导师：运怀立) - 《天津财经大学硕士论文》- 2021-05-01	0.6%（62） 是否引证：否
3	基于系统动力学-Visual Studio集成模型的区域碳排放中长期变化趋势预测——以四川省为例 李乔楚;陈军华;何京; - 《环境污染与防治》- 2022-12-15	0.6%（60） 是否引证：否
4	自适应裂变聚类算法的研究及应用 卢诗展(导师：程龙生) - 《南京理工大学博士论文》- 2020-10-01	0.5%（51） 是否引证：否
5	基于K-means聚类的煤炭港区TSP浓度变化及影响因素分析 林翔宇;张艳;封学军;林志端;朱信源;沈金星; - 《中国港湾建设》- 2022-07-25	0.5%（46） 是否引证：否
6	基于数据流的挖掘算法研究 陈照阳(导师：黄上腾) - 《上海交通大学硕士论文》- 2008-01-01	0.4%（43） 是否引证：否
7	我国温室气体监测技术应用及减排措施 束胜全;孙友文;徐亮;刘建国; - 《能源环境保护》- 2023-01-10 09:52	0.4%（39） 是否引证：否
8	长三角地区能源消费碳排放影响因素分解与预测研究 熊慧敏(导师：唐兆希) - 《浙江财经大学硕士论文》- 2019-12-01	0.4%（38） 是否引证：否
9	气候智慧型农业技术碳计量方法学初探 柏振忠;钟雨欣;胡婉玲;王红玲; - 《湖北农业科学》- 2022-12-25	0.3%（31） 是否引证：否
10	气候变化背景下韧性城市建设的意义与路径 孙永平;刘玲娜; - 《国家治理》- 2023-01-29	0.3%（31） 是否引证：否
11	基于Copula函数的北京市气候变化及人类活动对NDVI影响的识别 左斌斌;钟伟强;谭超;刘达;程涛; - 《水利与建筑工程学报》- 2023-02-15	0.3%（31） 是否引证：否

参赛队号：（参赛队无须填写，参赛队号由大赛官网自动生成）
2023年（第九届）全国大学生统计建模大赛参赛作品
参赛学校：
重庆大学
论文题目：
基于聚类和惩罚回归的四川省碳排放实证分析
参赛队员： 张轩铭陈勇杭王晓雪
指导老师： 徐建文

基于聚类和惩罚回归的四川省碳排放实证分析

摘要

在中国式现代化的新时代背景下，能源转型和绿色低碳发展已经成为国家和地方政策的关注领域。在这些政策的驱动

参赛学校:	重庆大学
论文题目:	基于聚类和惩罚回归的四川省碳排放实证分析
参赛队员:	张轩铭陈勇杭王晓雪
指导老师:	徐建文

下，四川省积极探索能源转型和碳排放降低的道路，并研究行业能源结构差异，协同推进绿色低碳发展。

为了更好地探究四川省行业能源结构的差异，本研究采用DBSCAN聚类方法，对不同行业的能源结构数据进行聚类并采用SSE和SC作为评估指标进行聚类效果评估，并采用SVM（二次）验证聚类的鲁棒性，最后利用惩罚回归模型对四川省的碳排放进行了核算。

首先，采用四川省能源结构数据，运用DBSCAN聚类方法对数据进行聚类，同时使用SSE和SC两种评估指标对聚类结果进行了评估分析。结果表明，四川省的能源结构存在行业差异，随着行业的不同，能源结构的组成表现出多样性，得出了最优聚类结果，即聚类数为16个。然后，在利用聚类结果进行二次SVM验证时，结果表明，四川省不同行业的能源结构相对于随机聚类得到的样本更具有可分性和集中性。这表明，DBSCAN聚类得到的簇对于SVM模型的监督学习具有很好的鲁棒性。

最后，本研究将聚类结果应用于惩罚回归方法中，对四川省的碳排放进行了核算。利用不同的正则化系数，对聚类结果依次进行了岭回归，Lasso回归，Elastic Net回归。研究结果表明，弹性网络惩罚回归模型具有最好的性能，能够很好地预测碳排放情况。这些结果为四川省环保政策的制定提供了关键信息，同时为实现可持续发展提供了理论和实践支持。

总之，在中国式现代化的时代背景下，本研究揭示了四川省行业能源结构的差异，探讨了行业能源结构的转型路径和相应的碳排放核算方法。这些研究成果对于推动四川省的绿色低碳发展至关重要，同时也提供了其他地区类似的研究方向和方法，为实现中国的绿色低碳目标提供了指导意义。

关键词：碳排放；中国式现代化；DBSCAN；惩罚回归

目录

表格与插图清单

基于聚类和惩罚回归的四川省碳排放实证分析

一、 引言

（一）研究背景和意义

当今气候变化正给自然界和人类社会造成广泛而普遍的影响，联合国政府间气候变化专门委员会（IPCC）第六次评估报告指出，通过排放温室气体，人类活动已毋庸置疑引起了全球变暖。报告对工业革命以来的气候变化进行了归因，发现自然因子和气候系统的内部变率对一百多年来气候变化的贡献几乎为零，换言之，火山爆发和太阳活动对工业化以来的增暖影响几乎可以忽略不计，而2011年至2022年全球地表温度比1850年至1900年高出1.1℃，而在近期全球升温可能达1.5℃，可以说是人类活动造成了1850年至今的几乎所有全球升温。

图 1 全球碳排放量分布图

由全球碳排放量分布图可以发现，碳排放的密集区域主要在中国，西欧，北美等国家和地区，中国作为世界上最大的发展中国家和经济体，消费碳排放占到了总排放量的三分之一。近年来，中国政府出台了大量的法规和政策以促进低碳经济的发展，并承担起了节能减排的重任[1]。西部地区的可持续发展是国家发展战略的重要内容，由于经济条件和技术水平限制，西部地区的环境承载力接近上限，四川省作为中国西南地区的经济中心，处于工业化、城镇化加速期，需要加快脚步进行经济转型升级。因此，把握四川省碳排放因素的特征，对于减少西南地区区域间碳排放的政策制定和实施具有重要意义。

（二）国内外研究现状

随着经济增长，在石油危机爆发后，能源稀缺、极端天气引起了全球的学者对碳排放因素的研究。

在1989年Kaya恒等式在IPCC研讨会上提出，通过链式乘积形式分解出多种影响因素，将人口、人均国内生产总值、能源强度及单位能耗的排放量相联系。

1994年，通过对IPAT模型的改进，STIRPAT模型[8]是多变量随机回归非线性模型，考虑人口、财富、技术等因素的变动对环境的影响，还允许根据研究对象进行添加、修改或分解相关影响因素。

20世纪末，Ang[9]等在Divisia分解法上加以改进，形成对数平均分解法LMDI，目前是研究碳排放变化和能源消耗的常用方法，消除了Divisia分解法的残差项，解决了数据收集中的零值和负值问题。

马江（2017）运用排放系数法和投入产出法[2]对四川省1995年至2016年的居民消费直接碳排放和间接碳排放进行核算和测量，为区域的政策制定提供理论依据。陈军华[3]（2021）研究测算了四川省2000-2018年15种能源消费的碳排放量，并使用LMDI分解方法将碳排放量增量分解为不同的影响因素。

（三）研究目的和方法

尽管国内外对于碳排放影响因素研究已经有较多文献，模型和方法应用都较为广泛，但因素选择的主观性在各模型中似乎是不可避免的，如LMDI分解法分解结果往往具有不唯一性，可能造成结果的不确定性。STIRPAT模型无法区分变量之间的因果关系，且变量之间常有较强的共线性问题。

本研究旨在探究四川省的碳排放特征、影响因素和规律，为制定有效的碳排放控制政策提供理论依据。首先，本研究将基于数据特征进行聚类，消除主观因素对分类的影响，从分类结果反推研究内在的实际意义。其次，结合四川省的实际情况，采用惩罚回归方法对共线性问题进行处理，并回归分析碳排放因素的影响。最终，通过研究结果，为中国式现代化发展中的西部地区提供碳排放控制政策，促进该地区可持续发展。本研究的意义在于解决现有研究中因素选择的主观性问题，并提供一种新的研究方法，为推动碳排放控制和可持续发展提供参考。

二、研究方法

（一）碳排放核算方法

碳排放核算已成为面临全球气候变化问题的重要任务之一。本文数据是采用的IPCC部门核算法[4]，这是一种广泛接受和使用的国际性碳排放核算方法，并可适用于许多行业和活动，中国碳核算数据库对四川省的各部门进行了碳排放核算。IPCC核算法包括确定排放源和活动数据采集、排放因子确定、碳排放量计算、数据分析和结果报告等四个阶段。其计算公式如下，其中，是指j部门的化石燃料i产生的二氧化碳排放；代表对应的化石燃料种类和部门产生的化石燃料消耗量；指燃烧每一单位化石燃料产生的热值；(碳含量)对于化石燃料i而言每一对应净热值的排放量；称作氧合效率，是指化石燃料燃烧时候的氧化率。

（二）DBSCAN聚类算法

DBSCAN算法[14]是一种基于密度聚类的算法，对于处理不规则形状的簇具有很好的效果。它可以有效地过滤噪声点，不需要预先指定簇的数量，通过最小点数和邻域半径两个参数来确定簇的大小和边界，进而将距离比较近的点组成一个密集区域。

对于给定的数据集，任取数据集Data中的两点，邻域参数(,MinPts)定义如下：

定义1：是样本的半径，表示以为圆心，为领域半径，定义的-邻域的圆形区域范围。

定义2：MinPts是样本的区域密度阈值，可判断是否作为核心点。当样本的-领域范围邻居样本值大于区域密度阈值MinPts，确定为核心点。

对于DBSCAN聚类算法，除领域参数定义外，常用的重要概念还有以下几个：

(1)核心对象(core point)：对于一个点，如果其某个半径r内包含的点数大于等于参数MinPts，则称为一个核心对象。核心对象是DBSCAN算法中的关键，因为它们可以组成簇。

(2)直接密度可达(directly density-reachable)：对于两个样本点，如果的邻域内的点且为核心对象，则被称为的直接密度可达点。

(3) 密度可达（density-reachable）：对于两个样本点和，如果存在一个样本点序列的直接密度可达点，则称为的密度可达点。

(4) 密度相连（density-connected）：对于两个样本点和，如果它们存在一个样本点使得和都是的密度可达点，则称和是密度相连的。

图 2 DBSCAN算法示意图

DBSCAN算法通过一组领域参数(,MinPts)确定数据集中样本分布的紧密情况，算法中的一个聚类簇由给定数据集中任意一个核心对象唯一确定，当空间中某区域中的样本对象数量超过给定参数MinPts的值，可遍历该区域对象直接生成聚类簇。

碳排放的复杂性决定了我们需要更加深入地研究它们的产生机制和影响因素。DBSCAN算法作为一种基于密度的聚类算法，可以发现任意形状的聚类簇，并且不需要提前指定聚类的数量。

（三）惩罚回归模型理论和应用

惩罚回归是一种统计建模方法，用于在线性回归模型中引入惩罚项，以控制模型的复杂度并提高其泛化能力。它在解决高维数据和多重共线性等问题时特别有用。惩罚项通常基于参数的大小进行调整，通过对参数进行惩罚，可以约束模型中参数的增长，从而避免过拟合。

惩罚回归基于最小化目标函数来估计模型参数。目标函数由两部分组成：损失函数和惩罚项。损失函数衡量了模型对训练数据的拟合程度，而惩罚项用于控制模型的复杂度。

惩罚回归有几种常见的模型建立方法，其中最常用的是岭回归[15]（Ridge Regression）和lasso回归（Lasso Regression）。下面对三种常见的惩罚回归原理进行阐述。

1.岭回归理论

岭回归通过添加L2正则化项来约束模型参数。L2正则化项是参数的平方和的乘以一个惩罚系数，它使得参数的值趋向于较小的范围。岭回归能够同时减小参数的估计值和方差，适用于存在多重共线性的情况。

损失函数：目标函数：2.Lasso回归理论

Lasso回归使用L1正则化项[17]，它是参数的绝对值之和的乘以一个惩罚系数。L1正则化具有稀疏性，即可以将某些参数估计为零，从而实现特征选择。Lasso回归在具有大量特征和需要特征选择的情况下很有用。

损失函数：目标函数：3.弹性网络回归理论

弹性网络回归（Elastic Net Regression）[18]是一种结合了岭回归和lasso回归的线性回归方法。在弹性网络回归中，目标函数由两部分组成：损失函数和惩罚项。惩罚项包括两个部分：L1正则化和L2正则化。L1正则化通过参数的绝对值之和来约束模型的复杂度，实现了特征选择的效果。L2正则化通过参数的平方和来限制参数的增长，减小参数估计的方差。

损失函数：目标函数：其中，是观测值的实际值，是模型的预测值，是回归系数，p是自变量的数量，是控制正则化程度的超参数。

总的来说，惩罚回归主要有三个优点，第一控制过拟合，惩罚回归通过引入惩罚项来控制模型的复杂度，减少了过拟合的风险，提高了模型的泛化能力；第二，处理高维数据：惩罚回归对于高维数据的建模效果较好，可以应对大量的自变量，并减少共线性带来的问题；第三，Lasso回归通过L1正则化实现特征选择，能够识别对碳排放影响最显著的特征，提供有关影响碳排放的关键因素的信息。

在消费碳排放研究中，惩罚回归能够提供对碳排放影响因素的量化评估，并揭示关键的影响因素。它可以帮助研究人员识别出重要的变量，并提供政策建议和行动指导，以减少个人或家庭的碳排放。然而，在应用惩罚回归时需要考虑到参数调节和估计偏差的问题，并结合领域知识进行解释和验证。

三、数据来源和预处理

（一）数据来源

本论文中二氧化碳排放数据的来源包括中国碳核算数据库和四川省统计年鉴[6]。其中，中国碳核算数据库提供了 2000-2019 年的二氧化碳排放数据[11][13]，数据的来源包括互联网调查以及国家和地方政府的统计数据等。而四川省统计年鉴提供了该省相关的经济和能源统计数据，包括工业、农业、交通运输、能源和环保等领域的数据。以中国碳核算数据库数据集为例，涉及17个能源种类，覆盖47种不同的社会生产经济部门，整理如表1，表2。

表 1 数据集能源种类

编号	能源种类	编号	能源种类
1	原煤	10	汽油
2	洗精煤	11	煤油
3	其他洗煤	12	柴油
4	煤砖	13	燃油
5	焦炭	14	液化石油气
6	煤气	15	炼厂气
7	可燃气	16	其他石油产品
8	焦化产品	17	天然气
9	原油		

编号能源种类编号能源种类
1 原煤 10 汽油
2 洗精煤 11 煤油
3 其他洗煤 12 柴油
4 煤砖 13 燃油
5 焦炭 14 液化石油气
6 煤气 15 炼厂气
7 可燃气 16 其他石油产品
8 焦化产品 17 天然气
9 原油

在社会经济部门中，除了行业领域编号，还进一步划分了产业类别。

表 2 数据集社会经济生产部门及类别

编号	社会经济部门	类别	
1	农、林、牧、渔、水利	第一产业	
2	煤炭开采与选矿	能源生产	制造业
3	石油和天然气开采		
4	黑色金属采矿和选矿		
5	有色金属开采与选矿		
6	非金属矿产开采与选矿		
7	其他矿产开采及选矿		
8	木材和竹子的伐木和运输		
9	食品加工		
10	粮食生产		
11	饮料生产		
12	烟草加工		
13	纺织工业		
14	服装和其他纤维制品	轻工制造	
15	皮革，毛皮，羽绒及相关产品		
16	木材加工，竹，甘蔗，棕榈纤维及稻草制品		
17	家具制造		
18	造纸及纸制品		
19	印刷和复制		
20	文化、教育和体育用品		
21	石油加工和焦化	能源生产	
22	化工原料及化工产品	重型制造业	
23	医疗及医药产品	轻工制造	
24	化学纤维	重型制造业	
25	橡胶制品		
26	塑料制品		
27	非金属矿产品		
28	黑色金属的冶炼和压制		
29	有色金属的冶炼和压制		
30	金属制品		
31	普通机械		
32	特殊用途设备		
33	运输设备		
34	电气设备及机械	高科技产业	
35	电子及电讯设备		
36	仪器，仪表，文化和办公机械		
37	其他制造业		
38	废料和废物		
39	电力、蒸汽、热水的生产和供应	能源生产	
40	天然气的生产和供应		
41	生产及供应自来水	重型制造业	

42	建设	建筑	
43	运输、仓储、邮电服务	服务行业	
44	批发、零售贸易及餐饮服务		
45	其他服务行业		
46	城市居民能源使用情况	家庭	
47	农村居民能源使用情况		

编号社会经济部门类别
1 农、林、牧、渔、水利第一产业
2 煤炭开采与选矿
能源生产
制造业
3 石油和天然气开采
4 黑色金属采矿和选矿
5 有色金属开采与选矿
6 非金属矿产开采与选矿
7 其他矿产开采及选矿
8 木材和竹子的伐木和运输
9 食品加工
10 粮食生产
11 饮料生产
12 烟草加工
13 纺织工业
14 服装和其他纤维制品轻工制造
15 皮革，毛皮，羽绒及相关产品
16 木材加工，竹，甘蔗，棕榈纤维及稻草制品
17 家具制造
18 造纸及纸制品
19 印刷和复制
20 文化、教育和体育用品
21 石油加工和焦化能源生产
22 化工原料及化工产品重型制造业
23 医疗及医药产品轻工制造
24 化学纤维
重型制造业
25 橡胶制品
26 塑料制品
27 非金属矿产品
28 黑色金属的冶炼和压制
29 有色金属的冶炼和压制
30 金属制品
31 普通机械
32 特殊用途设备
33 运输设备
34 电气设备及机械
高科技产业
35 电子及电讯设备
36 仪器，仪表，文化和办公机械
37 其他制造业
38 废料和废物
39 电力、蒸汽、热水的生产和供应能源生产
40 天然气的生产和供应
41 生产及供应自来水重型制造业
42 建设建筑
43 运输、仓储、邮电服务服务行业
44 批发、零售贸易及餐饮服务
45 其他服务行业
46 城市居民能源使用情况家庭
47 农村居民能源使用情况
在收集和使用这些数据的过程中，严格遵守了相关的法律法规和研究伦理，同时也保证了数据的隐私和保密性。
（二）数据预处理
在数据预处理中，编号16的其他石油产品在四川省能源消耗中均为零值，被去除，共计16种有效能源；与此同时，木材与

竹子的伐木与运输历年均为零值，不纳入考虑，总计有效经济部门为46个。为初步探究能源种类的不同差异情况，我们将2012-2019年8年中的能源使用情况可视化如下图。

图 3 2012年-2019年各能源的二氧化碳排放总量示意图

能源的使用量变化由折线图可得，原煤使用量呈下降趋势，而天然气使用量呈上升趋势。能源碳排放的堆积图显示，虽然原煤使用量减少，但仍占较大比重，四川省能源仍以煤炭为主。汽油、柴油和天然气使用量增加，意味着经济高速发展，各领域能源用量也随之增加。

四川省碳排放数据包含46个相关领域，能源差异性和变化趋势是重要特征。对各领域数据进行了最大最小标准化处理，以便比较能源使用差异。处理公式：通过处理，得到各个领域行业的归一化数据表，将原始数据与归一化后的数据进行整理，得到下表。

表 3 部分归一化数据对比

相关产业领域	原煤		精煤		煤砖		焦炭		汽油	
	原数据	归一化	原数据	归一化	原数据	归一化	原数据	归一化	原数据	归一化
金属	0.552	0.014	2.460	0.063	0.037	0.047	39.03	1.000	0.061	0.001
电力	29.10	1.000	0.015	0.000	3.224	0.110	0.026	0.001	0.127	0.004
食品	0.482	1.000	0.025	0.051	0.036	0.075	0.067	0.140	0.187	0.389
橡胶	0.026	0.582	0.001	0.003	0.000	0.012	0.000	0.020	0.045	1.000
家具	0.011	0.170	0.003	0.001	0.000	0.000	0.067	1.000	0.023	0.352

相关产业领域原煤精煤煤砖焦炭汽油

原数据归一化原数据归一化原数据归一化原数据归一化原数据归一化

金属 0.552 0.014 2.460 0.063 0.037 0.047 39.03 1.000 0.061 0.001

电力 29.10 1.000 0.015 0.000 3.224 0.110 0.026 0.001 0.127 0.004

食品 0.482 1.000 0.025 0.051 0.036 0.075 0.067 0.140 0.187 0.389

橡胶 0.026 0.582 0.001 0.003 0.000 0.012 0.000 0.020 0.045 1.000

家具 0.011 0.170 0.003 0.001 0.000 0.000 0.067 1.000 0.023 0.352

四川省碳排放数据集中，不同行业的碳排放量存在巨大的绝对值差异。一些行业在焦煤的排放量高达39.03Mt，而相较于其他行业，最高的汽油仅有0.045Mt，最大最小标准化能更好地展示数据间精细的差异和变化趋势。例如黑色金属冶炼中焦煤的比重最大，食品加工中原煤的用量最高，被较好地呈现。

a.原始数据 b.归一化结果

图 4 各部门能源消耗热力图

从图中可以清晰的看到，相同数量的分级之下，由于原始数据绝对大小带来的影响，大多数行业领域的能源种类区分度并不大，色块基本以黄色为主，而右图中对比归一化之后的热力图，消除了偏离的绝对数据带来的掩盖效应，各领域的能源层次更加丰富精细，特征也更加明显。

四、聚类分析

基于上述标准化处理的数据，我们进行DBSCAN聚类，并结合轮廓系数（SC）和簇内平方误差和（SSE）进行综合评价。SC是用来衡量聚类结果的好坏的指标，其值越接近1表示聚类结果越好。SSE用来衡量聚类结果中簇内样本点的相似度，其值越小表示聚类效果越好。

轮廓系数[16]是指对于第i个样本点，设它所属的簇为，则它到簇内的其它所有样本点的平均距离为，设它到簇的其它所有样本点的平均距离为，则该样本点的轮廓系数可以定义为：

其中，表示该样本点所属的簇集合中与该样本点距离平均值最小的簇的平均距离，即：表示点到点之间的距离。而表示该样本点与同一个簇内其他样本点的平均距离：而簇内平方误差[5]是每个簇所有样本点距离该点质心距离的平方和，计算公式为，其中，n表示样本数目，k表示簇数目，表示第i个样本点，表示第j个簇质心，表示第i个样本点属于第j个簇的权重（1表示属于该簇，0表示不属于）。

因此借助归一化后的行业能耗矩阵，利用MATLAB代码遍历寻找最优SC和SSE，得到当C=16，SC=0.6，SSE=5,结果表明簇间差异明显，簇内的近似程度较高，聚类分析结果较好。

考虑到DBSCAN的无监督性，为了进一步验证聚类结果的鲁棒性，我们又引入了不同年份的归一化矩阵结合混淆矩阵和ROC曲线对结果进行验证。

图 5 SVM[7]（二次）混淆矩阵和ROC

可以发现，混淆矩阵真实类正确率除孤立类之外，准确率平均在93.7%，且ROC曲线大部分都趋近于1,因此我们认为该分类在能源特征层面上是合理的，样本具有可分性和集中性，并整理结果如下表。

表 4 聚类结果

社会经济部门	分类	行业类别	主要能源
农、林、牧、渔、水利	1	第一产业	
其他矿产开采及选矿		能源生产	
建设		建筑	柴油\汽油
批发、零售贸易及餐饮服务		服务行业	
其他服务业			
煤炭开采与选矿	2	能源生产	
非金属矿产开采与选矿			
食品加工		轻工制造	
饮料生产			原煤\天然气
纺织工业			
木材加工，竹，甘蔗，棕榈纤维及稻草制品			

造纸及纸制品			
其他制造业		高科技产业	
电力、蒸汽、热水的生产和供应		能源生产	
石油和天然气开采	3	能源生产	炼厂气\天然气
黑色金属采矿和选矿	4	能源生产	其它气体\焦炭
有色金属开采与选矿	5	能源生产	其它气体\汽油
粮食生产	6	轻工制造	
皮革，毛皮，羽绒及相关产品			
家具制造			
化学纤维		重型制造业	天然气
金属制品			
特殊用途设备			
电气设备及机械		高科技产业	
天然气的生产和供应		能源生产	
烟草加工	7	轻工制造	其它洗煤
服装和其他纤维制品	8	轻工制造	
印刷和记录介质复制			
橡胶制品			
塑料制品		重型制造业	
电子及电讯设备			汽油\原煤\天然气
仪器，仪表，文化和办公机械		高科技产业	
生产及供应自来水		重型制造业	
城市			
农村		家庭	
文化、教育和体育用品	9	轻工制造	
黑色金属的冶炼和压制			
普通机械		重型制造业	
废料和废物		高科技产业	
石油加工和焦化	10	能源生产	清洁煤\其它洗煤
化工原料及化工产品	11	重型制造业	天然气
医疗及医药产品	12	轻工制造	其它洗煤\原煤\汽油
非金属矿产品	13	重型制造业	进程
有色金属的冶炼和压制	14	重型制造业	天然气\焦炭\其他气体
运输设备	15	重型制造业	天然气\其它洗煤\汽油
运输、仓储、邮电服务	16	服务行业	煤油\柴油\汽油

社会经济部门分类行业类别主要能源
农、林、牧、渔、水利 1
第一产业
柴油\汽油
其他矿产开采及选矿能源生产
建设建筑
批发、零售贸易及餐饮服务服务行业
其他服务业
煤炭开采与选矿 2 能源生产
原煤\天然气
非金属矿产开采与选矿
食品加工
轻工制造
饮料生产
纺织工业
木材加工，竹，甘蔗，棕榈纤维及稻草制品
造纸及纸制品
其他制造业高科技产业
电力、蒸汽、热水的生产和供应能源生产
石油和天然气开采 3 能源生产炼厂气\天然气
黑色金属采矿和选矿 4 能源生产其它气体\焦炭
有色金属开采与选矿 5 能源生产其它气体\汽油
粮食生产
6
轻工制造
天然气
皮革，毛皮，羽绒及相关产品
家具制造
化学纤维

重型制造业
金属制品
特殊用途设备
电气设备及机械高科技产业
天然气的生产和供应能源生产
烟草加工 7 轻工制造其它洗煤
服装和其他纤维制品 8
轻工制造
汽油\原煤\天然气
印刷和记录介质复制
橡胶制品
塑料制品重型制造业
电子及电讯设备
仪器，仪表，文化和办公机械高科技产业
生产及供应自来水重型制造业
城市
农村家庭
文化、教育和体育用品
9 轻工制造
焦炭
黑色金属的冶炼和压制
普通机械重型制造业
废料和废物高科技产业
石油加工和焦化 10 能源生产清洁煤\其它洗煤
化工原料及化工产品 11 重型制造业天然气
医疗及医药产品 12 轻工制造其它洗煤\原煤\汽油
非金属矿产品 13 重型制造业进程
有色金属的冶炼和压制 14 重型制造业天然气\焦炭\其他气体
运输设备 15 重型制造业天然气\其它洗煤\汽油
运输、仓储、邮电服务 16 服务行业煤油\柴油\汽油
在此分类基础上，我们对不同簇类的数据特征进行提取，计算了对应的均值，方差，中位数，并给出了常见的分位数。
表 5 簇类特征

类别	总和	均值	方差	最小值	P0.25	中位数	P0.75	最大值
1	23.51	1.57	11.24	0	0	0.00	1.23	12.91
2	67.25	4.48	99.43	0.0048	0.198	0.63	3.13	40.41
3	5.77	0.38	0.37	0	0	0.00	1.20	1.85
4	3.60	0.24	0.22	0	0	0.02	0.33	1.64
5	1.08	0.07	0.02	0	0	0.00	0.15	0.43
6	3.86	0.26	0.20	0	0.002	0.03	0.24	1.63
7	0.07	0.00	0.00	0	0	0.00	0.01	0.04
8	24.17	1.61	11.72	0	0.004	0.01	0.41	10.96
9	66.22	4.41	99.29	0	0.049	1.05	3.48	40.56
10	18.13	1.21	6.36	0	0	0.05	0.64	8.11
11	10.59	0.71	0.91	0	0	0.36	1.06	3.76
12	0.52	0.03	0.00	0	0	0.00	0.06	0.17
13	20.15	1.34	3.06	0	0.012	0.45	1.87	4.51
14	2.62	0.17	0.05	0	0	0.03	0.42	0.69
15	1.48	0.10	0.02	0	0	0.03	0.21	0.39
16	27.60	1.84	11.56	0	0	0.00	3.00	8.99

类别总和均值方差最小值 P0.25 中位数 P0.75 最大值
1 23.51 1.57 11.24 0 0 0.00 1.23 12.91
2 67.25 4.48 99.43 0.0048 0.198 0.63 3.13 40.41
3 5.77 0.38 0.37 0 0 0.00 1.20 1.85
4 3.60 0.24 0.22 0 0 0.02 0.33 1.64
5 1.08 0.07 0.02 0 0 0.00 0.15 0.43
6 3.86 0.26 0.20 0 0.002 0.03 0.24 1.63
7 0.07 0.00 0.00 0 0 0.00 0.01 0.04
8 24.17 1.61 11.72 0 0.004 0.01 0.41 10.96
9 66.22 4.41 99.29 0 0.049 1.05 3.48 40.56
10 18.13 1.21 6.36 0 0 0.05 0.64 8.11
11 10.59 0.71 0.91 0 0 0.36 1.06 3.76
12 0.52 0.03 0.00 0 0 0.00 0.06 0.17

13 20.15 1.34 3.06 0 0.012 0.45 1.87 4.51
14 2.62 0.17 0.05 0 0 0.03 0.42 0.69
15 1.48 0.10 0.02 0 0 0.03 0.21 0.39
16 27.60 1.84 11.56 0 0 0.00 3.00 8.99

将聚类结果数据特征可视化，得到不同簇类的能源消耗堆积图和箱线图，可以发现在碳排放总量方面，第二簇和第九簇体量相当，但能源结构差异显著，观察箱线图，绝大部分碳排放值都集中在区间内，离群点同样出现在第二簇和第九簇，这意味着四川省碳排放的产业结构调整需要重点从这两方着手。

图 6 能源分布图

五、惩罚回归分析

（一）回归模型建立

STIRPAT（Stochastic Impacts by Regression on Population, Affluence, and Technology）模型是一种常用的环境影响评估模型，它通过回归分析来探究人口、富裕度和技术对环境影响的关系。STIRPAT模型中的指标通常量纲不同，但在本研究数据集中对各行业的碳排放进行了测算，量纲均为，且聚类种类多，为此回归模型借助STIRPAT进行修改，具体公式如下，其中为常数项，e为误差项，代表聚类种类，指回归系数，n代表聚类簇数。

对各类数据取对数，部分结果展示如下表，完整数据在附录。

表 6 部分数据对数处理结果

年份 类别	2000	2001	2002	2003	2004	2005
1	1.72265	1.784575	1.823502	2.036376	2.242504	2.24204
2	3.820479	3.834683	4.03818	4.302104	4.381153	4.397185
3	-0.08105	-0.02259	0.050675	0.035983	0.139035	0.061531
4	-1.53239	-1.49975	-1.42292	-1.31853	-1.11337	-1.13641
5	-2.41151	-2.43257	-2.34743	-2.3089	-2.17434	-2.19876

年份
类别 2000 2001 2002 2003 2004 2005
1 1.72265 1.784575 1.823502 2.036376 2.242504 2.24204
2 3.820479 3.834683 4.03818 4.302104 4.381153 4.397185
3 -0.08105 -0.02259 0.050675 0.035983 0.139035 0.061531
4 -1.53239 -1.49975 -1.42292 -1.31853 -1.11337 -1.13641
5 -2.41151 -2.43257 -2.34743 -2.3089 -2.17434 -2.19876

在惩罚回归建立之前，先用最小二乘法和方差膨胀因子VIF对簇类的共线性进行模拟检验。

反映了该自变量与其他自变量的相关性程度，其取值范围为[1, +∞)，通常认为当时存在较严重的多重共线性问题。其中，表示第k个自变量的方差膨胀因子，表示将第k个自变量作为因变量，其他自变量作为自变量，所建立的回归模型的决定系数。结果显示在最小二乘法回归中的情况下，VIF值偏大，落在(98.4,+∞]中，因此进行惩罚回归对共线性进行处理是合理的。

为了评估惩罚回归效果，引入两类指标进行评估：一类是用于评估模型预测性能的指标，另一类是用于评估模型特征选择效果的指标。

预测性能指标主要包括均方误差（Mean Squared Error，MSE）和R平方Coefficient of Determination，）。

均方误差是用来衡量预测值与实际值之间差异的指标，MSE越小，则预测值与实际值越相似，模型的预测能力越好，其值为非负；R平方是用来衡量回归模型对实际值解释能力的指标，越接近1，则说明模型对实际值的解释能力越好，预测结果越准确。其中，n表示样本数，表示第i个样本的实际值，表示第i个样本的预测值，表示实际值的平均值。

评估模型特征选择效果的指标为稀疏度s，指的是参数向量或特征向量中非零元素的比例，即保留了多少有用的信息。增加正则化项可以使得参数或特征向量更易取得零元素从而达到稀疏的目的。其中，s示稀疏度，n表示样本数，p表示参数或特征的个数，表示指示函数，当参数或特征的系数不等于零时为1，反之为0。稀疏度s的取值范围在之间，表示的是参数或特征中非零元素所占的比例。当参数或特征向量完全稠密时，s=1；当参数或特征向量完全稀疏时，s=0。

应用MATLAB程序，对2000年-2014年15年的碳排放数据进行惩罚回归训练，在岭回归中，我们通过增加系数λ，来对模型的系数进行惩罚,控制模型的复杂度和泛化能力,正则化系数越大，模型的复杂度越小，对训练数据的拟合程度也越小，但是泛化能力越强；反之亦然。因此我们对正则化系数进行不断迭代，确认最大的决定系数。

图 7 R平方迭代图

设置λ迭代步长为0.01，可以发现，在时，能取到的最大的是0.74691，而Mse是0.49342，拟合的效果并不理想。进一步探究发现，岭回归结果的稀疏度s为1，主要原因是参数维度过高，岭回归无法较好的处理。

图 8 岭回归拟合效果图

再观察预测值和实际值，以为例，可知约为，约为，预测差值在，偏离了27.67%。

进而我们用Lasso 回归对数据进行降维，增加一个正则化项λ，使得模型系数的 L1 范数总和小于一个固定值，从而使一些系数为 0，达到选择变量的目的，同时减少了模型的过拟合程度。并通过Lasso拟合系数轨迹图，综合预测性能和解释性能寻找最优特征系数点。

图 9 Lasso系数拟合轨迹图

Lasso 回归系数轨迹图展示不同正则化系数 Lambda 下各个特征的系数大小，横轴是λ值，纵轴是各个特征的系数,选择通过交叉验证得到最小误差的 Lambda值作为最优正则化系数，从而得到最优的模型。随着正则化系数 Lambda的增加，系数曲线逐渐稳定并出现水平段，水平段之前的系数为非零，由图可知非零项共计7项，为0.9991,Mse为,整理如下表，

表 7 Lasso回归结果

系数	a1	a2	a3	a4	a5	a6	a7	coef
取值	0.2823	0.1203	0.1660	0.0200	0.0850	0.0005	0.3461	2.8986
均方误差 Mse		1.4774×10-4		决定系数 R2		0.9991		

系数 a1 a2 a3 a4 a5 a6 a7 coef

取值 0.2823 0.1203 0.1660 0.0200 0.0850 0.0005 0.3461 2.8986

均方误差

Mse

1.4774×10-4 决定系数

R2

0.9991

最终确认非零系数有7项，代入到回归方程中得到，带入数据求解，观察拟合效果，得到实际观测值和预测值的拟合效果图。以以为例，可知约为，约为，预测差值在，误差仅在0.54%，比岭回归预测有显著提升。

图 10 Lasso拟合效果图

但参数被缩减到7个，稀疏度s为0.4375，这是相对较高的稀疏度，意味参数拥有相对较少信息，可能会有过拟合和欠拟合情况，为此我们进一步用Elastic Net回归模型对碳排放进行建模和预测。

通过调整L1正则化（Lasso）和L2正则化（岭回归）之间的权衡参数alpha，。使用交叉验证等方法对建立的Elastic Net回归模型进行评估和优化，我们找到了最佳的模型复杂度，此时的值为0.9999,MSE为。

带入回归模型进行验证，得到如下的拟合效果图，可以看到，Elastic Net的直线性良好。

图 11 Elastic Net拟合效果图

（二）回归模型综合分析

将上述三种回归结果汇总，得到系数及评价指标综合表。

表 8 系数及评价指标综合表

类别	岭回归		Lasso		Elastic Net	
a0	4.6178	Coef0	0	Coef0	0	Coef0
1	0.0390		0	2.8986	0.0236	2.3273
2	0.0628	R2	0.2823	R2	0.4319	R2
3	0.0159	0.74691	0	0.9991	0.0669	0.9994
4	0.0139	Mse	0	Mse	0	Mse
5	0.0125	0.49342	0	1.4774×10-4	0	1.8309×10-5
6	0.0334		0		0	
7	0.0342	s	0	s	0.0108	s
8	0.0321	1	0.1203	0.4375	0.1132	0.75
9	0.0341	λ	0.1660	λ	0.1534	λ1=λ2
10	0.0268	2.35	0.0200	0.0081	0.0540	2.7826×10-4
11	0.0072		0		0.0058	
12	-0.0070	α	0	α	0	α
13	0.0256		0.0850		0.0538	0.5
14	0.0292		0.0005		0.0797	
15	0.0359		0.3461		0.0368	
16	0.0344		0		0.0047	

类别岭回归 Lasso Elastic Net

a0 4.6178 Coef0 0 Coef0 0 Coef0

1 0.0390 0 2.8986 0.0236 2.3273

2 0.0628 R2 0.2823 R2 0.4319 R2

3 0.0159 0.74691 0 0.9991 0.0669 0.9994

4 0.0139 Mse 0 Mse 0 Mse

5 0.0125 0.49342 0 1.4774×10-4 0 1.8309×10-5

6 0.0334 0 0

7 0.0342 s 0 s 0.0108 s

8 0.0321 1 0.1203 0.4375 0.1132 0.75

9 0.0341 λ 0.1660 λ 0.1534 λ1=λ2

10 0.0268 2.35 0.0200 0.0081 0.0540 2.7826×10-4
11 0.0072 0 0.0058
12 -0.0070 α 0 α 0 α
13 0.0256 0.0850 0.0538 0.5
14 0.0292 0.0005 0.0797
15 0.0359 0.3461 0.0368
16 0.0344 0 0.0047

根据系数的大小来评估各个变量对碳排放的相对重要性，系数的绝对值越大，说明对碳排放的影响越大。可以发现，三种回归系数中，第二簇的系数基本占比是最大的,为此我们进一步分析第二簇分类特征。

表 9 第二簇特征分析

社会经济部门	簇类	行业类别	主要能源	占比
煤炭开采与选矿		能源生产	原煤	22.49%
非金属矿产开采与选矿		能源生产	原煤	1.2%
食品加工		轻工制造	原煤\天然气	1.5%
饮料生产		轻工制造	原煤	1.5%
纺织工业	2	轻工制造	原煤\天然气	1.5%
木材加工，竹，甘蔗，棕榈纤维及稻草制品		轻工制造	原煤\天然气\汽油	0.19%
造纸及纸制品		轻工制造	原煤	1.69%
其他制造业		高科技产业	原煤	0.15%
电力、蒸汽、热水的生产和供应		能源生产	原煤	69.63%

社会经济部门簇类
行业类别主要能源占比
煤炭开采与选矿能源生产原煤 22.49%
非金属矿产开采与选矿能源生产原煤 1.2%
食品加工轻工制造原煤\天然气 1.5%
饮料生产轻工制造原煤 1.5%
纺织工业 2
轻工制造原煤\天然气 1.5%
木材加工，竹，甘蔗，棕榈纤维及稻草制品轻工制造原煤\天然气\汽油 0.19%
造纸及纸制品轻工制造原煤 1.69%
其他制造业高科技产业原煤 0.15%
电力、蒸汽、热水的生产和供应能源生产原煤 69.63%

结合行业能源分布表和饼状图可以发现，第二簇行业横跨三个领域，包括能源生产、轻工制造和高科技产业，能源供应主要是以原煤为主。

和其他化石燃料相比，原煤能量密度低，燃烧效率低，约为30%左右，并且在煤炭开采与选矿领域所得的原煤又反哺给该行业进行进一步开发，与此同时，大部分都集中在电力、热水供应上。究其原因，四川省经济发展较为快速，加上地理位置限制，四川省地处西南内陆，地势复杂，交通不便，电力输送困难，因此在能源消费方面，该省较为依赖煤炭资源，原煤消耗相对较高。

图 12 第二簇碳排放能源分布图

为了对碳排放进行预测，我们需要选择最好的回归模型，将拟合的到的数据取e的指数，结合实际的碳排放总量，绘制在同一张历年碳排放图中。可以发现，由于指数的特性，尽管岭回归预测结果取ln之后相差不大，但在实际的碳排放量中，随着时间推移误差越来越大。而Lasso和Elastic Net对于实际碳排放仍有极好的拟合，但在2014年末端Lasso存在部分偏移。

图 13 三种惩罚回归拟合效果图

在一定超参数范围内，岭回归能够在一定程度上提高模型的泛化能力和可解释性，然而，岭回归方法对于权重较小或零的特征无法直接消除且导致特征缩放问题；LASSO回归能够显著提高模型的解释性和稀疏性，同时仍保持较高预测准确性，然而，LASSO回归方法忽略了弱相关特征的问题；Elastic Net回归能够综合利用岭回归和LASSO回归的优点，并取得较好的折衷结果。

综上所述，从超参数调节和稀疏度方面来看，Elastic Net回归方法表现最佳，其次为LASSO回归，岭回归方法的调节和稀疏度表现相对较弱，基础的最小二乘法不能满足我们需要的要求。

（三）模型预测

应用Elastic Net回归结果，对2015年—2019年数据进行预测，并将真实值与预测值及相应的差值整理如下表，

表 10 Elastic预测

年份 数据	2015	2016	2017	2018	2019
真实值	5.8063	5.7655	5.7647	5.6914	5.7531
预测值	5.8284	5.7107	5.7012	5.5943	5.6615
差值	-0.0221	0.0548	0.0635	0.0972	0.0916

年份
数据 2015 2016 2017 2018 2019
真实值 5.8063 5.7655 5.7647 5.6914 5.7531
预测值 5.8284 5.7107 5.7012 5.5943 5.6615

差值 -0.0221 0.0548 0.0635 0.0972 0.0916

(1)平均误差：计算这五组误差的平均值为0.0570。平均误差接近零，预测值与实际值整体上趋于一致.

(2)方差分析：计算这五组误差的方差为0.0023，较小的方差表示预测值相对稳定。

(3)趋势分析：观察这五组误差的趋势，即误差是逐渐增加还是逐渐减少。误差呈现递增，可能存在系统性的偏差。

图 14 误差趋势图

将这五组误差作为模型的残差，观察残差的分布情况和统计特征，绘制残差的直方图、散点图等图表，检查是否存在明显的模式或异常值。

图 15 残差分析图

可以看出这里的误差在2018年有明显升高。呈现反弹上升趋势，进一步深入研究数据发现，能源消耗的意外增长似乎与天气影响有关：2018年和2019年，四川省连续发生暴雨洪灾，导致部分地区洪水泛滥、山体滑坡；2018年和2019年，部分地区出现严重的高温天气，2021年遂宁市出现高温天气，局部气温突破40℃；2020年3月，四川省南充市出现冰雹天气，造成农作物损失和房屋损毁；2020年至2022年，四川省均发生多起严重的森林火灾。研究表明，极端天气事件对电力需求的影响比一般天气事件更显著，而且高温天气对电力需求的影响更加显著，在四川省的研究中，夏季高温天气对城市电力需求的影响更为显著。

六、政策建议

综合研究结果，在聚类的过程中，可以发现第二类和第八类的碳排放总量处于高位，对回归模型的系数进行分析，可以看出第二类（原煤）、第八类（汽油）、第九类的系数值较大，因此我们具体分析了一下这几个类之间与行业的联系。原煤、电力、热水供应多（类别2）：原煤的燃烧以及电力和热水的产生通常依赖于化石燃料，例如煤炭和天然气。四川省处于西南地区，交通不便，电力运输困难，常住人口数量逐年递增，对于电力的要求负担不断提升，也因此需要加大原煤燃烧发电的需要，这些过程会产生大量二氧化碳（CO2）等温室气体的排放。

针对这种情况，可以采取以下政策措施：

推广可再生能源：增加可再生能源如太阳能、风能和水能的利用，减少对化石燃料的依赖。

提高能源效率：通过技术改进和能源管理措施，减少能源消耗，降低碳排放。

加强能源结构转型：逐步减少对高碳能源的依赖，鼓励清洁能源的开发和使用。

汽油（类别8）：汽油的燃烧主要由机动车辆引起，包括城市和农村地区的汽车使用，因此汽油的量主要与城市农村居民消耗有着密切关系。在分析原煤燃烧中提到四川省常住人口数量逐年递增，这也反映了四川省还正处于一个经济的上升期，随着人口的增长，对于汽车的保有量也逐年递增。而汽车尾气排放作为主要的碳排放来源之一。

针对这种情况，可以采取以下政策措施：

提倡公共交通和非机动交通：加强公共交通系统的建设，鼓励市民使用公共交通工具，减少个人汽车使用。同时，鼓励步行、骑行等非机动交通方式。

推广电动车和混合动力车辆：促进电动车和混合动力车辆的推广和普及，减少传统汽车尾气排放。

焦炭（类别9）：焦炭涉及高温工业生产，通常需要大量的能源消耗，而黑色金属冶炼作为高温工业的代表，与焦炭的燃烧有着密切关系。黑色金属冶炼就包括化石燃料的燃烧。这导致了显著的碳排放。四川省正处于经济发展期，对于工业发展迫切需要，通过聚类分析的结果，我们可以看出焦炭的量占比相较于高位也有明显改善，可以看出工业在不断转型。

针对这种情况，可以采取以下政策措施：

推广清洁生产技术：鼓励使用更清洁、高效的冶炼技术，减少能源消耗和碳排放。

强化排放控制：对焦炭和黑色金属冶炼企业实施严格的排放控制措施，包括污染物排放限制和排放监测。

七、结论

首先，本文对于碳排放研究领域中的主观性判断和多重共线性情况进行了调查，发现聚类算法和惩罚回归能分别解决主观问题和多重共线性，为此我们展开了实际的数据分析和研究。

对四川省46个行业利用能源种类占比不同的特征进行聚类，得到16个分类簇，从能源特征角度对行业情况进行分析，为碳排放实证研究带来新视角。

进一步采取三种惩罚回归模型，借助机器学习，训练模型，探究16类对四川省碳排放的实际影响情况，最终了解到第二簇类对碳排放影响最大，其次为第八第九簇类，分别主要能源种类为原煤、汽油和焦炭。背后原因主要依次为，第一，四川省地处西南，电力运输不便，依赖大量原煤进行日常和工业电力、热水等供应；第二，四川省的人口数量和经济发展相对较快，车辆保有量日益增加，这也导致了汽油用量的增加；第三，钢铁工业是焦炭用量最大的行业之一，它是钢铁炼制过程的精髓所在，同时也是四川省重点发展的行业之一，因此其排放量相对较多。

进一步选用Elastic Net回归模型进行预测，仍然得到较好的拟合结果，这一点很好的解决了碳排放研究领域中，模型过拟合的情况，进一步丰富碳核算，碳减排相关领域的理论体系和实践效能。

总之，本文在方法、数据和结论等方面经过了较为全面的考虑、研究和分析。其结论不仅能够为相关领域提供一定价值的实践指导和理论参考，同时也为今后的研究提供了有益的启迪和方法借鉴。

参考文献

[1] 鲍健强,苗阳 & 陈锋.(2008).低碳经济:人类经济发展方式的新变革. 中国工业经济(04),153-160. doi:10.19581/j.cnki.ciejournal.2008.04.018.

[2] 马江. (2021). 四川省居民消费碳排放测算及特征分析. (2017-3), 89-95.

[3] 陈军华, & 李乔楚. (2021). 成渝双城经济圈建设背景下四川省能源消费碳排放影响因素研究--基于lmdi模型视角. 生态经济, 37(12), 7.

[4] 赵磊, 陈德珍, 刘光宇, 栾健, & Thomas H.Christensen. (2010). 垃圾热化学转化利用过程中碳排放的两种计算方法. 环境科学学报, 30(8), 1634-1641.

[5] 吴军. 数学之美[M]. 人民邮电出版社, 2013.

[6] 四川省政府. 四川省统计年鉴[M]. [四川省政府, 2022.

[7] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1):11.

[8] York, R. , Rosa, E. A. , & Dietz, T. . (2003). Stirpat, ipat and impact: analytic tools for unpacking the driving forces of environmental impacts. Ecological economics(3), 46.
[9] Ang, B. W. , Huang, H. C. , & Mu, A. R. . (2009). Properties and linkages of some index decomposition analysis methods. Energy Policy, 37(11), 4624-4632.
[10] Shan, Y. , Guan, D. , Zheng, H. , Ou, J. , Li, Y. , & Meng, J. , et al. (2018). Data Descriptor: China CO2 emission accounts 1997-2015.
[11] Yuli Shan, Qi Huang, Dabo Guan, & Klaus Hubacek. (2020). China co2 emission accounts 2016–2017. Scientific Data, 7(1).
[12] Assessment to china's recent emission pattern shifts. Earth's Future.
[13] Shan, Y. , Liu, J. , Liu, Z. , Xu, X. , Shao, S. , & Wang, P. , et al. (2016). New provincial co2 emission inventories in china based on apparent energy consumption data and updated emission factors. Applied Energy, 184(DEC.15), 742-750.
[14] Birant, D. , & Kut, A. . (2007). St-dbscan: an algorithm for clustering spatial–temporal data. Data & Knowledge Engineering, 60(1), 208-221.
[15] Hoerl, A. E. , & Kennard, R. W. . (2000). Ridge regression: biased estimation for nonorthogonal problems. Technometrics A Journal of Stats for the Physical Chemical & Engineering ences, 42.
[16] Rousseeuw, P.J. (1987). “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”, Journal of Computational and Applied Mathematics.
[17] Hui Z . Taylor & Francis Online :: The Adaptive Lasso and Its Oracle Properties - Journal of the American Statistical Association - Volume 101, Issue 476[J]. Journal of the American Statistical Association, 2006, 101(476):1418-1429.
[18] Zou H , Hastie T . Addendum: "Regularization and variable selection via the elastic net'' [J. R. Stat. Soc. Ser. B Stat. Methodol. 67 (2005), no. 2, 301–320; MR2137327].[J]. journal of the royal statistical society, 2010, 67(5):768-768.

附录

致谢

非常荣幸能够在此向您呈上我们的论文，并表达我们最诚挚的谢意和感激之情。在完成《基于聚类和惩罚回归的四川省碳排放实证分析》的研究过程中，得到了许多人的支持和帮助，在此我们要向他们表示衷心的感谢。

首先，我们要衷心感谢我的导师徐老师。徐老师在整个研究过程中给予了我们悉心的指导和宝贵的建议。他的悉心指导和启发性的思路使我能够顺利地完成研究工作，并在学术上有了新的突破和进步。同时，徐老师对我们的关怀和支持也使我们倍感温暖和鼓舞。

此外，我们还要感谢各位老师和同学们。感谢他们提供的良好学习环境和研究条件，为我们提供了宝贵的资源和支持。在实浓厚的学术氛围中，我们受到了他们的激励和启发

特别感谢参与本研究的调查对象和数据提供者，他们的积极参与和数据提供为研究结果的可靠性和准确性提供了重要依据。

最后，我们要感谢我的家人和朋友们。感谢他们一直以来的理解、支持和鼓励。他们给予我们精神上的支持和关心，使我们能够专注于研究工作，克服了各种困难和挑战。

再次衷心感谢所有支持和帮助过我们的人，正是有了你们的支持和鼓励，我们才能够顺利完成这篇论文的撰写。希望我们的研究成果能够对相关领域的发展和应用有所贡献，回报社会。同时，我们也将倍加珍惜这次学术研究的机会，为科学研究事业贡献自己的力量。

再次衷心致谢！

谨向各位致以最诚挚的问候和祝福！

2023年5月26日