

基于随机森林算法的辽宁省建筑业碳排放影响因素重要度研究

孔凡文¹,张贵恒¹,王英华²

(1.沈阳建筑大学管理学院,辽宁 沈阳 110168;2.沈阳城市建设学院管理学院,辽宁 沈阳 110167)

摘要:根据联合国政府间气候变化专门委员会(The Intergovernmental Panel on Climate Change, IPCC)提出的碳排放核算方法,基于全生命周期法建立了建筑业碳排放核算模型。通过收集整理建筑业碳排放相关影响因素,引入随机森林算法,并以2006—2020年辽宁省建筑业相关数据进行实例分析,得出了近年来在辽宁省建筑业碳排放的各项影响因素中,重要度较高的为人均生产总值、施工面积、从业人数、建筑业人均生产总值、能源强度和人口数量。最后,结合2006—2020年辽宁省建筑业碳排放变化情况及关键影响因素对建筑业提出了减排建议。

关键词:建筑业;碳排放;随机森林算法;影响因素

中图分类号:X322;TU111.195

文献标志码:A

近年来,由于二氧化碳排放量快速升高导致的全球变暖对人类日常生活的影响愈发严重,气候问题逐渐成为最受全球人民关注的问题之一。中国作为全球节能减排活动的坚定参与者,提出将采取强有力的措施加强对二氧化碳排放的管控。张涑贤等^[1]和黄振华^[2]采用可拓展的随机性的环境影响评估模型(Stochastic Impacts by Regression on Population, Affluence, and Technology, STIRPAT)模型,分别对陕西省和重庆市建筑业碳排放的影响因素进行了实证分析,并在此基础上对建筑业的节能减排工作提出了建议;李爽等^[3]利用扩展的Kaya恒等式分解出具体指标,以定量分析各解释变量的同时,也发现各变量均对建筑业碳排放产生了正向影响;冯博等^[4]通过Kaya恒等式分解指标,分析了中国各省建筑业碳排放与经济

形势的脱钩现象;蒋博雅等^[5]基于对数平均迪氏指数分解法(Log-Mean Divisia Index Method, LMDI)对江苏省建筑业碳排放影响因素进行了解析,通过解析各因子在江苏省建筑业碳排放中的贡献率来剖析其驱动效应。

综合来看,现有关于建筑业碳排放影响因素的研究热度较高,但在影响因素的选择上大多是基于分析模型的固有因素,而关于影响因素重要度的研究较少。因此,笔者以2006—2020年辽宁省建筑业相关数据为例,核算并分析碳排放现状,梳理建筑业碳排放相关因素,并引入随机森林算法分析各项影响因素的重要程度,进而识别其中的关键因素,以为后续节能减排工作的开展提供参考。

一、建筑业碳排放核算模型及随机森林算法

1. 建筑业碳排放核算范围界定

对于碳排放的核算方法,主要有投入产出法、实测法、碳排放因子法以及全生命周期法。其中,全生命周期法是以各项活动作为划分阶段的依据来对某一特定生产过程的碳排放全过程进行链条式碳排放核算。中国尚未对建筑业碳排放进行统计,考虑到核算投入与产出所需使用的投入产出表在中国的更新次数较少,而实测法主要应用于微观或单体的建筑碳排放核算,因此笔者采用使用较为广泛的全生命周期法结合碳排放因子法对辽宁省建筑业碳排放量进行核算。

根据全生命周期法对辽宁省建筑业碳排放核算划定范围,可将建筑业的生命周期划分为建材生产及运输、建筑施工、既有建筑运行及建筑拆除 4 个阶段。建材生产及运输阶段的碳排放主要来自钢材、水泥、玻璃和铝材等建筑材料的生产及运输过程;建筑业在施工建造及拆除阶段主要的碳排放来自煤炭、汽油、柴油、热力以及电力等能源的使用;既有建筑运行阶段的碳排放主要来源于建筑采暖、照明、通风等过程中的能源消耗^[6],根据建筑类别及统计年鉴数据分类,可将既有建筑运行碳排放分为公共建筑运行碳排放、居住建筑运行碳排放及城镇采暖产生的碳排放 3 类,其中,公共建筑能耗主要由交通运输业、仓储及邮电通讯业、批发及零售贸易业、餐饮业和其他行业的能源消耗组成;居住建筑能源消耗主要来源于城镇生活能源消费和乡村生活能源消费;城镇采暖能源消耗主要来源于锅炉房供暖用煤和热电联产中用于建筑采暖的煤炭使用。

2. 建筑业碳排放核算模型

《1996 年 IPCC 国家温室气体清单指南修订本》提出了根据每种碳排放源活动及碳排放因子确定碳排放量的方法,这也是使用最为广泛的算法,其计算公式为

$$E=AD\times EF$$

(1)

式中: E 为碳排放量; AD 为活动水平; EF 为碳排放因子。

根据全生命周期法对建筑业划分的阶段可建立建筑业全生命周期碳排放核算模型,其碳排放量核算公式为

$$E_{CO_2}=E_e+E_m+E_p$$

(2)

式中: E_{CO_2} 为建筑业碳排放量; E_e 、 E_m 、 E_p 分别为建筑施工阶段、建材生产及运输阶段和建筑运行阶段的碳排放量。

建筑施工阶段碳排放量核算公式为

$$E_e=\sum_{i=1}^3C_i\times AQN_i\times\alpha_i+C_h\times\beta+C_e\times\gamma$$

(3)

式中: C_i 为第 i 种能源的消耗量; AQN_i 为第 i 种能源的平均低位发热值; α_i 为第 i 种能源的单位热值排放因子; C_h 为建筑业热力消耗; β 为热力碳排放因子; C_e 为建筑业电能消耗; γ 为电能碳排放因子。

建材生产及运输阶段碳排放量核算公式为

$$E_m=\sum_{j=1}^4C_j\times(\alpha_{1j}+\alpha_{2j}\times d_j)$$

(4)

式中: C_j 为第 j 种能源的消耗量; α_{1j} 为第 j 种建筑材料的碳排放因子; α_{2j} 为第 j 种建筑材料运输时的碳排放因子; d_j 为第 j 种建筑材料运输时的运输距离。

建筑运行阶段碳排放量核算公式为

$$E_p=\sum_{l=1}^3C_l\times AQN_l\times\alpha_l+\sum_{m=1}^2C_m\times AQN_m\times\alpha_m+\sum_{n=1}^2C_n\times AQN_n\times\alpha_n$$

(5)

式中: C_l 为公共建筑能耗中第 l 种行为的能源消耗量; AQN_l 为第 l 种行为消耗能源的平均低位发热值; α_l 为第 l 种行为消耗能源的单位热值排放因子; C_m 为居住建筑运行中第 m 种行为的能源消耗量; AQN_m 为第 m 种行为消耗能源的平均低位发热值; α_m 为第 m 种行为消耗能源的单位热值排放因子; C_n 为居住建筑运行中第 n 种行为的能源消耗量; AQN_n 为

第 n 种行为消耗能源的平均低位发热值; α_n 为第 n 种行为消耗能源的单位热值排放因子。

3. 建筑业碳排放影响因素选取

在对人类活动造成环境影响的研究中, IPAT (Imact-Population-Affluence-Technology) 是较为常见的研究模型。IPAT 模型认为环境受到人口数量、人均财富和科学技术水平这 3 个主要因素的影响。人口数量的增长、物质生活水平的提高和过度的资源开发会导致环境压力的不断增加, 人类活动对环境的影响是人口增长、经济增长与科技进步共同作用的结果。在此基础上, 选取出符合建筑业特征的相关指标, 同时结合对相关参考文献^[7-14]的梳理, 笔者选取的影响因素为城镇化率、能源强度、人均生产总值、人口数量、建筑业发展水平、建筑业人均生产总值、施工面积、从业人数、建筑业产值以及劳动生产率, 并统计了出现的频次(见图 1), 其中, 城镇化率为城镇人口与总人口的比值, 能源强度为能源消耗与建筑业产值的比值, 建筑业发展水平为建筑业产值与总产值的比值, 劳动生产率为建筑业增加值与从业人数的比值。

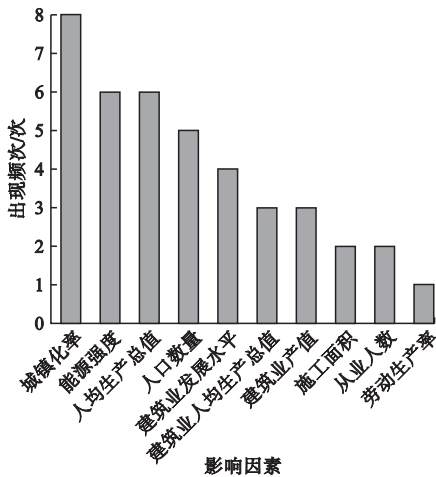


图 1 影响因素频次统计

4. 随机森林算法

随机森林(Random Forest)是一种以决策树(Decision Tree)为基础的监督式集成学习算法,可用于分析复杂的相互作用特征,在处理噪声数据或数据缺失时具有很好的鲁棒性,其集成算法的特性能够提高计算准确性。加之样本随机抽取和特征随机选择这两项随

机性的引入也使得随机森林算法不易产生过拟合问题。由于建筑业碳排放影响因素的数量多、联系紧密,且彼此之间相互影响,因此使用随机森林算法计算所选定影响因素的重要度。

随机森林算法实现重要度计算的途径是以基尼系数作为衡量系标,用 GI 表示基尼系数,基尼系数的计算公式为

$$GI_m = \sum_{k=1}^K p_{mk} (1 - p_{mk}) = 1 - \sum_{k=1}^K p_{mk}^2 \quad (6)$$

式中: GI_m 为数据在节点 m 的基尼系数; K 为数据中有 K 个类别; p_{mk} 为在节点 m 上类别 k 所占比例。

假设有 r 个特征 X_1, X_2, \dots, X_r , 则特征 X_j 在节点 m 的重要度 $VIM_{jm}^{(Gini)}$, 即节点 m 分支前后基尼指数变化量为

$$VIM_{jm}^{(Gini)} = GI_m - GI_l - GI_r \quad (7)$$

式中: VIM 为变量重要度; GI_l 和 GI_r 分别为分支后 2 个新节点的基尼系数。

若特征 X_j 在随机森林中某一决策树 i 上出现的节点为集合 M , 则特征 X_j 在第 i 棵树的的重要度 $VIM_{ij}^{(Gini)}$ 为

$$VIM_{ij}^{(Gini)} = \sum_{m \in M} VIM_{jm}^{(Gini)} \quad (8)$$

假设随机森林中有 n 棵树, 则特征 X_j 的重要度 $VIM_j^{(Gini)}$ 为

$$VIM_j^{(Gini)} = \sum_{i=1}^n VIM_{ij}^{(Gini)} \quad (9)$$

最后, 将所得重要度评分进行归一化处理, 得到各项影响因素的重要度 VIM_j 为

$$VIM_j = \frac{VIM_j^{(Gini)}}{\sum_{j=1}^r VIM_j^{(Gini)}} \quad (10)$$

二、实例分析

1. 辽宁省建筑业碳排放核算

根据式(2)对 2006—2020 年辽宁省建筑业碳排放进行核算, 其计算结果如图 2 所示。从碳排放总量上看, 2006—2020 年间辽宁省建筑业碳排放量的变化趋势呈现“几”字形。2006—2010 年期间, 辽宁省建筑业碳排放以年均 26.33% 的增速增长; 2011 年增

速高达 64.33%;2011—2014 年期间其排放量稳定在 $3.0 \times 10^8 \text{t}$;2015 年的排放量以较大幅度下降;后期其排放量逐年降低,仅在 2020 年出现小幅上升。从碳排放的分布结构看,在辽宁省建筑业碳排放中,2006—2020 年,建筑运行阶段和建材生产及运输阶段的碳排放占总排放量的 98%~99%,施工及拆除阶段仅占其中的 1%~2%(见图 3)。

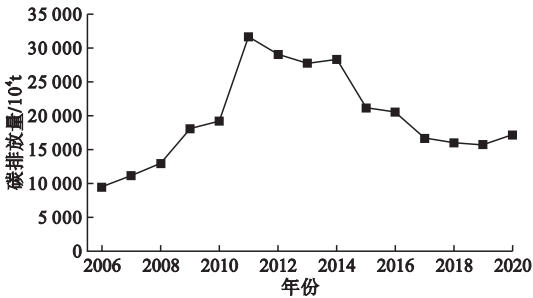


图 2 2006—2020 年辽宁省建筑业碳排放量

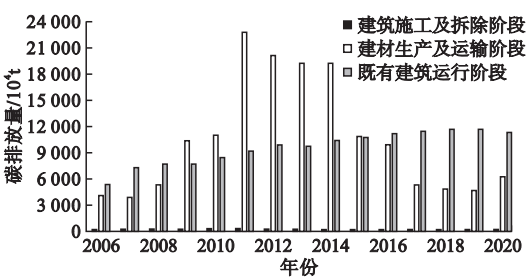


图 3 2006—2020 年辽宁省建筑业分阶段碳排放量

2. 影响因素重要度计算及分析

将以上统计总结得到的影响因素及核算所得的 2006—2020 年辽宁省建筑业碳排放量进行整理(见表 1)。在进行重要度分析时,由于某些影响因素有非唯一的计量单位,例如,表 1 中人均生产总值这一项既可以以亿元/人作为单位,也可以以百万元/人作为单位,此时就会出现因计量单位不同导致的数字大小问题,可能对后续分析产生影响,因

表 1 2006—2020 年辽宁省建筑业碳排放量及影响因素

年份	碳排放量/ 10^4 t	城镇化率/%	能源强度/ $10^8 (\text{J} \cdot \text{万元}^{-1})$	人均生产总值/万元	人口数量/万人	建筑业发展水平/%	建筑业人均生产总值/万元	施工面积/ 万 m^2	从业人数/万人	建筑业总产值/亿元	建筑业劳动生产率/万元/人
2006	9 373.84	59.84	0.49	1.99	4 210.40	21.16	17.81	12 102.30	99.65	1 775.00	13.66
2007	11 212.65	60.13	0.48	2.43	4 231.70	20.40	21.14	13 704.30	99.35	2 100.00	15.32
2008	12 907.75	61.02	0.44	2.86	4 246.10	20.64	22.92	14 616.80	109.31	2 505.20	16.11
2009	18 097.13	61.24	0.37	3.02	4 256.00	26.35	25.52	18 760.50	132.61	3 384.60	17.37
2010	19 206.56	63.95	0.30	3.27	4 251.70	33.75	28.29	26 806.96	165.80	4 690.30	17.87
2011	31 635.12	65.98	0.25	3.84	4 255.00	38.02	36.22	34 906.30	171.70	6 218.30	31.45
2012	29 117.04	67.88	0.18	4.20	4 244.80	42.26	37.12	40 040.90	203.19	7 543.30	35.89
2013	27 715.53	68.83	0.16	4.53	4 238.00	44.93	43.61	42 289.00	197.88	8 629.70	32.44
2014	28 276.37	69.38	0.14	4.72	4 244.20	39.21	45.00	47 861.00	174.45	7 851.10	31.68
2015	21 115.61	66.90	0.19	4.78	4 229.70	26.79	40.05	28 937.10	135.18	5 413.80	30.09
2016	20 540.74	69.69	0.25	4.82	4 232.00	19.26	30.92	20 390.70	127.02	3 927.00	33.12
2017	16 611.01	70.26	0.26	5.17	4 196.50	17.00	35.16	16 506.00	104.89	3 688.30	35.72
2018	15 949.60	70.72	0.26	5.61	4 191.90	15.01	47.34	13 659.80	74.54	3 528.40	30.12
2019	15 740.98	70.72	0.25	5.94	4 190.20	14.27	52.90	15 312.80	67.20	3 554.60	32.19
2020	17 073.62	73.73	0.22	6.03	4 165.90	15.19	48.19	16 235.00	79.17	3 815.30	31.18

注:能源及建材使用《辽宁省建筑业统计年鉴》相关数据;建材碳排放因子、能源单位热值碳排放因子及运输相关参数来自《建筑碳排放计算标准》(GB/T51366—2019);能源低位发热值、电力能源碳排放因子及热力系数来自《综合能耗计算通则》(GB/T2589—2020);相关影响因素数据均来自《辽宁省建筑业统计年鉴》。

此需要对其进行处理,但处理的前提是不能失去数字的相对意义,即处理前数字越大代表人均生产总值越高,处理后的数据也不能失去这个特性,因此采用归一化的方法对各项数据进行处理,归一化的计算公式为

$$e = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{11}$$

式中: e 为归一化的计算结果; X 为各项影响

因素的具体数据值; X_{\max} 为 X 所代表的影响因素中的最大值; X_{\min} 为 X 所代表的影响因素中的最小值。

在利用随机森林算法进行影响因素重要度研究时,最重要的是对决策树的数量(ntree)进行确定,以降低误差或避免过拟合现象的发生。通过调用 `sklearn` 中的 `GridSearchCV` 函数,利用交叉验证的方式,

对 ntree 进行调整,得到 ntree 与该算法准确度之间的关系(见图4),可以看出当 ntree 处于 35 ~ 55 时,该算法的准确率较高,且当 ntree 为 35 时,其准确率最高,达到了 83.1%,以此确定模型 ntree 为 35。通过对表 1 的统计结果进行归一化处理,在 python3.8 中调用 sklearn 中的 Random Forest 函数,可实现基于随机森林算法对各项影响因素重要度的计算(见表2)。

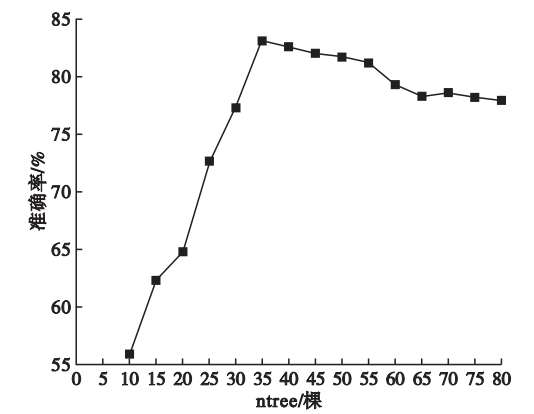


图4 GridSearchCV 计算准确率

表2 辽宁省建筑业各影响因素的重要度

指标	重要度
人均生产总值	0.137 643
施工面积	0.137 355
从业人数	0.122 763
建筑业人均生产总值	0.121 849
能源强度	0.119 893
人口数量	0.106 944
城镇化率	0.070 694
建筑业发展水平	0.066 718
建筑业总产值	0.060 377
建筑业劳动生产率	0.055 765

从计算结果可以看出,重要度较高的影响因素依次为人均生产总值、施工面积、从业人数、建筑业人均生产总值、能源强度以及人口数量,而城镇化率、建筑业发展水平、建筑业总产值和建筑业劳动生产率对辽宁省建筑业碳排放的影响则相对较小。

据此可得出以下结论:①在后续的发展中,首先应当关注经济对建筑业节能减排的作用。在辽宁省大力推进城镇化建设的建筑业发展初期,由于追求经济的快速发展,导致了碳排放量较高;随着经济的进一步发展,辽

宁省通过提高建筑业发展水平和转变建筑业发展模式来对建筑业碳排放实现控制,逐步发挥出经济发展对碳排放量的抑制作用。②施工面积和从业人数的重要度较高,说明辽宁省应当提高对建筑业中绿色设计及施工的重视程度,同时强化从业人员在各项生产活动中的低碳意识。③建筑业人均生产总值与能源强度的重要度较高,说明辽宁省应提高对建筑业相关技术升级的重视程度。前者与建筑业碳排放在变化趋势方面的相似表明建筑业作为高碳排放行业,其单位劳动力在创造较高产业价值的同时也造成了较高的碳排放,侧面印证了当前辽宁省建筑业在生产阶段的技术有待提高;能源强度能够反映单位能源消耗的产出价值,能源强度的降低意味着产出同等价值时消耗的能源量更少,而技术升级是提高能源利用效率的重要途径。④居民在日常生活中所进行的照明、取暖、炊事等行为,都会造成二氧化碳的排放。人口数量的重要度较高,说明辽宁省一方面应加强对居民关于节约能源的引导,强化居民的节能意识;另一方面,也可通过提供补贴或降低价格等方式,鼓励居民更多地购置更节能的家用电器。

三、结 语

通过对辽宁省建筑业碳排放进行模型测算,同时引入随机森林算法对建筑业碳排放影响因素进行影响程度高低的分析,筛选出了建筑业碳排放的重要影响因素,为二氧化碳的减排措施及政策制定提供了一定的参考。同时,研究发现,在以往所选取的建筑业碳排放影响因素中,较少考虑的影响因素如施工面积和从业人数,在以辽宁省建筑业为研究对象的计算中却显现了较高的影响度。因此在后续关于建筑业碳排放的研究中,不妨拓宽思路,根据实际情况将更多影响因素纳入研究范围,再对各项影响因素进行综合考量,分析各项影响因素的影响机理及重要度,进而规划出更为合理的建筑业节能减排方案,以期助力中国早日实现“碳达峰、碳中和”的目标。

参考文献:

[1] 张涑贤,张丹. 基于 STIRPAT 模型的陕西省建筑业能源消耗影响因素研究[J]. 建筑经济, 2014(3):89-93.

[2] 黄振华. 基于 STIRPAT 模型的重庆市建筑碳排放影响因素研究[J]. 项目管理技术,2018, 16(5):55-60.

[3] 李爽,陶东,夏青. 基于扩展 STIRPAT 模型的我国建筑业碳排放影响因素研究[J]. 管理现代化,2017,37(3):96-98.

[4] 冯博,王雪青. 中国各省建筑业碳排放脱钩及影响因素研究[J]. 中国人口·资源与环境, 2015,25(4):28-34.

[5] 蒋博雅,黄宝麟,张宏. 基于 LMDI 模型的江苏省建筑业碳排放影响因素研究[J]. 环境科学与技术,2021,44(10):202-212.

[6] 周琳,宋梦歌. 沈阳市既有建筑碳排放量核算及节能减排对策研究[J]. 沈阳建筑大学学报(社会科学版),2017,19(5):509-513.

[7] 刘明达,蒙古军,刘碧寒. 国内外碳排放核算方法研究进展[J]. 热带地理,2014,34(2): 248-258.

[8] 唐赛,付杰文,武俊丽. 中国典型城市碳排放影响因素分析[J]. 统计与决策,2021,37 (23):59-63.

[9] 王幼松,石惠萌,闫辉. 基于 SDM 的中国建筑业碳排放强度分布及空间效应分析[J]. 工程管理学报,2021,35(5):1-6.

[10] 闫辉,刘惠艳,邱若琳. 基于逐步回归的建筑业碳排放影响因素分析[J]. 工程管理学报, 2021,35(2):16-21.

[11] 杨艳芳,李慧凤,郑海霞. 北京市建筑碳排放影响因素研究[J]. 生态经济,2016,32(1): 72-75.

[12] 褚智亮,杨永标,王旭东. 基于 STIRPAT 模型驱动建筑能耗增长影响因素的研究[J]. 电力与能源,2015,36(2):251-255.

[13] 邓小乐,孙慧. 基于 STIRPAT 模型的西北五省区碳排放峰值预测研究[J]. 生态经济,2016, 32(9):36-41.

[14] 祁神军,张云波. 中国建筑业碳排放的影响因素分解及减排策略研究[J]. 软科学,2013,27 (6):39-43.

The Importance Research of Influencing Factors of Carbon Emissions in Construction Industry in Liaoning Province Based on Random Forest Algorithm

KONG Fanwen¹,ZHANG Guiheng¹,WANG Yinghua²

(1. School of Management, Shenyang Jianzhu University, Shenyang 110168, China;2. School of Management, Shenyang Urban Construction University, Shenyang 110167, China)

Abstract: According to the carbon emission accounting method proposed by IPCC, the carbon emission accounting model of construction industry is established based on the whole life cycle method. By collecting and sorting out the influencing factors of construction industry carbon emissions, introducing random forest algorithm, and taking the relevant data of construction industry in Liaoning Province from 2006 to 2020 as an example, it is concluded that among the influencing factors of construction industry carbon emissions in Liaoning Province in recent years, important influencing factors are per capita GDP, construction area, number of employees, per capita GDP of construction industry, energy intensity and population. Finally, based on the changes of carbon emissions from construction industry in Liaoning Province from 2006 to 2020 and the key influencing factors, emission reduction suggestions for the construction industry are proposed.

Key words: construction industry; carbon emission; random forest algorithm; influencing factors
(责任编辑:徐聿聪 英文审校:林 昊)