

基于机器学习的建材消耗预测研究

卢 圣, 彭雨熙

广西财经学院管理科学与工程学院, 广西 南宁

收稿日期: 2024年4月18日; 录用日期: 2024年5月8日; 发布日期: 2024年5月17日

摘 要

建筑业作为国民经济支柱性产业, 在推动社会发展的同时也消耗了大量建材。我国在推动建筑行业高质量发展, 同时进行节约材料和节能减排的行动, 这对实现我国“双碳”目标具有重要意义, 也节约大量经济成本和资源。优化建筑建材消耗的一个重要手段是建筑建材精细化管理, 而进行这一工作的基础就是对建筑建材进行准确的预测, 从而支撑建筑运行优化管理, 实现节约建材的目标。本项目利用机器学习方法分析和研究建筑建材历史数据和影响因素, 构建建筑建材消耗预测体系, 以便预测建筑在设计、建造和使用阶段的建材需求。本项目旨在为建筑业的智能建造和可持续发展提供理论依据和实践指导。

关键词

机器学习, 建筑建材消耗, 预测模型

Research on Building Materials Consumption Prediction Based on Machine Learning

Sheng Lu, Yuxi Peng

School of Management Science and Engineering, Guangxi University of Finance and Economics, Nanning Guangxi

Received: Apr. 18th, 2024; accepted: May. 8th, 2024; published: May. 17th, 2024

Abstract

As a pillar industry of the national economy, the construction industry consumes a lot of building materials while promoting social development. China is promoting the high-quality development of the construction industry while saving materials and energy saving and emission reduction ac-

tions, which is of great significance to achieve China's "double carbon" goal, but also save a lot of economic costs and resources. An important means to optimize the consumption of building materials is the fine management of building materials, and the basis of this work is to accurately forecast building materials, to support the optimal management of building operations, and to achieve the goal of saving building materials. This project uses a machine learning method to analyze and study the historical data and influencing factors of building materials and builds a building materials consumption prediction system to predict the building materials demand in the design, construction and use stages of buildings. This project aims to provide a theoretical basis and practical guidance for intelligent construction and sustainable development in the construction industry.

Keywords

Machine Learning, Construction Building Materials Consumption, Predictive Modelling

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 建筑建材消耗预测的背景和方法

本研究将通过建筑的历史数据进行分析和处理, 以及考虑影响建筑建材消耗的因素, 构建建筑建材消耗预测影响因子。运用机器学习算法对建筑历史数据进行训练和拟合, 通过数据特征的相关性筛选出相关的特征之间的关系, 因特征的相关性有差异, 所以通过 *Lasso* 回归方法对原始特征进行进一步筛选, 只保留重要的特征。基于灰色预测对小样本数据集的优良性状, 得出各特征的预测值, 而后构建 SVR 预测模型, 将两者结合对建筑建材进行预测。

2013 年, 孙霞在《建筑节能与绿色施工管理的分析》中提出在当前能源需求庞大与供给不足的态势下, 建筑企业的施工管理加入绿色施工管理是建筑行业可持续发展的必由之路[1]。2015 年, 阮鹏在《建设工程绿色施工管理研究》中通过对研究对象的实证分析, 提出了绿色施工的技术措施、管理方法和现场经验, 在绿色施工的过程中采取具体的技术措施与管理方法, 以提供我国绿色施工管理工作的进程[2]。2016 年, 赵小荣在《建筑施工管理与绿色建筑施工管理》中提出建筑企业如何通过建筑施工的有效管理, 在兼具企业利润的同时, 实施绿色建筑施工管理, 是建筑企业长远发展的一项生存机制[3]。2017 年, 钱仁兴在《基于绿色施工理念下建筑施工管理探析》中提出在建筑施工中的绿色施工理念及践行原则, 并且从强化施工人员管理、建筑材料管理、施工设备管理及质量控制等方面, 阐述了基于绿色施工理念下新时期建筑施工管理的构建策略[4]。2019 年, 刘贵中在《绿色建筑施工管理在建筑施工管理中的应用探讨》中提出了从建筑的规范设计、施工及运行整个生命周期内贯穿绿色理念, 全面提高建筑整个过程中的环保和节约意识, 通过实施绿色建筑施工管理, 实现能源的节约, 这样不仅有利于保护环境, 而且对建筑行业的可持续发展也具有积极的促进作用[5]。

综上所述, 通过对建筑历史数据的分析和处理, 以及对影响建筑建材消耗的因素进行考虑, 我们构建了建筑建材消耗预测影响因子, 并运用机器学习算法对数据进行训练和拟合。通过对数据特征的相关性筛选和 *Lasso* 回归方法的进一步筛选, 我们成功地保留了重要的特征, 提高了预测的准确性和实用性。此外, 我们结合了灰色预测和 SVR 预测模型, 从而更加全面地预测了建筑建材的消耗情况。这些不仅为建筑业的资源规划和成本控制提供了有力支持, 也为未来建材消耗预测的研究和实践指明了方向。

2. 分析建筑的建材消耗数据特征的相关性

为了更好地理解建材消耗数据的特征及其之间的关系,我们采用了 *Pearson* 相关系数作为分析工具,以量化建材消耗数据特征之间的相关性。*Pearson* 相关系数是一种常用的统计方法,用于衡量两个变量之间的线性相关程度。其取值范围为[-1, 1], 其中 1 表示完全正相关, -1 表示完全负相关, 0 表示无相关性。通过计算 *Pearson* 相关系数, 我们可以评估建材消耗数据特征之间的相关性, 并据此做出相关性大小的判断。

在本研究中, 我们将首先收集并清理建筑建材消耗数据, 包括建材种类、消耗量、时间等信息。然后, 我们利用这些数据特征之间的 *Pearson* 相关系数, 根据相关系数的大小来判断它们之间的相关性。通过分析, 我们可以更深入地理解建材消耗数据的特征, 为后续的建材消耗预测模型构建提供重要的参考和依据。

我们收集并分析以下建筑建材的使用量数据: 内外墙涂料、电线和电缆、钢筋、石膏板、混凝土、地砖和墙砖、铁丝、钉子、螺丝螺帽、木模板、铝模板、砌块、钢管和方木、水泥以及防腐油、切割片、氧气和乙炔。这些数据将成为我们研究的主要对象, 通过对它们的收集、整理和分析, 我们将能够更好地了解建筑施工过程中各种建材的消耗情况, 为建材消耗的预测建立模型提供可靠的数据基础。结合

$$Pearson \text{ 相关系数公式 } r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}}, \text{ 而后通过各特征之间的相关性进行分析,}$$

得出各特征之间的热力图, 如图 1。

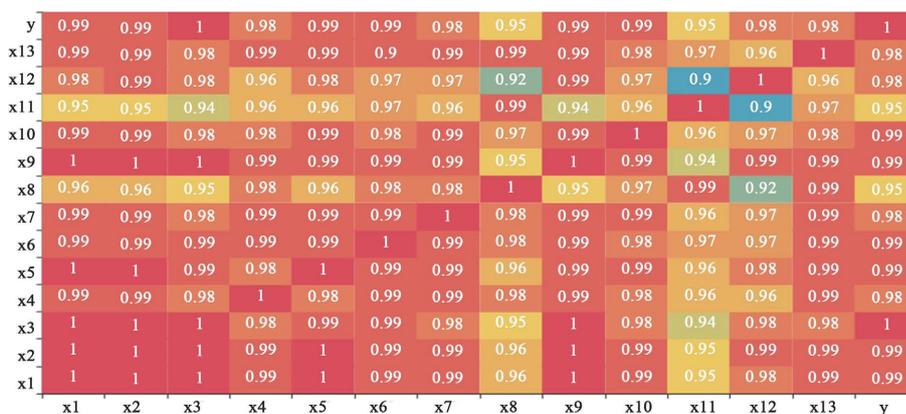


Figure 1. Heat map of correlation between features

图 1. 各特征之间的相关性热力图

由图 1 可知, 特征之间的相关系数以热力图的形式展示。热力图的数值表示了数值所在行特征和列特征的相关性, 数值越大相关性越大。当数值为 1 时特征之间完全相关, 当数值为 0 时特征之间完全无关, 相关系数越接近 1, 热力图对应位置的颜色越接近红色。

数据中的全部特征与建筑材料实际使用量(y)之间都存在高度的正相关关系, 按相关性大小排列, 依次是 x3、x1、x2、x5、x6、x9、x10、x4、x7、x12、x13、x8 和 x11。同时, 各特征之间存在着严重的多重共线性。例如, x1 (内外墙涂料)与其他全部特征的相关性都超过了 0.94。其中 x8 (木模板)和 x11 (钢管和方木)与其他特征的共线性较低, 可以考虑在后续的特征筛选中保留这两个特征。除此之外, x1 与 x2、x2 与 x3、x1 与 x3 等多对特征之间存在完全的共线性。

综上, 数据的全部特征都可以用于建筑材料实际使用量预测分析, 但特征之间存在着信息的重复, 需要对特征进行进一步筛选。

3. 使用 Lasso 回归选取建筑的建材消耗预测的关键特征

在本研究中我们将运用计算机工具结合 Lasso 回归方法, 旨在探究建筑建材消耗预测的关键特征选择。Lasso 回归(Least Absolute Shrinkage and Selection Operator)是一种经典的线性回归技术, 其特点在于引入 L1 正则化项以实现特征选择和模型的稀疏性。该方法通过对模型系数施加惩罚, 有效地压缩不重要特征的系数至零, 从而实现了特征选择的目的。Lasso 回归能够自动选择出重要特征, 实现模型的简化和解释性提升。

Lasso 参数估计公式 $\hat{\beta}(lasso) = \operatorname{argmin} \left\| y - \sum_{i=1}^p x_i \beta_i \right\|^2 + \lambda \sum_{i=1}^p |\beta_i|$ 。将数据投入公式中模拟结果, 我们将选取系数非零的特征作为提取的关键特征, 得出系数非零的特征有 x3、x5、x8、x10、x11、x12。

4. 使用灰色预测算法和 SVR 算法构建建筑的建材消耗预测模型

本研究的关键特征基于灰色预测对小样本数据集的优良性能, 首先对单个特征建立灰色预测模型, 得到各特征的预测值。然后对特征的训练集建立 SVR 预测模型, 将建立好的模型与灰色预测模型相结合, 使用灰色预测算法和对未来的建筑建材消耗进行预测, 如表 1。

Table 1. The grey prediction model predicts the values of the selected features

表 1. 灰色预测模型预测出所选特征的值

	X3	X5	X8	X10	X11	X12
预测值 1	11,980,347	10,500,423	73,394,636	40,107,127	1678	50,843,506
预测值 2	12,605,948	11,056,891	7,819,751	44,828,110	1774	57,905,229

Table 2. P-value and C-value returned by the grey prediction function to determine the prediction accuracy level of the screened features

表 2. 灰色预测函数返回的 P 值和 C 值对所筛选特征的预测精度等级判定

	P	C	预测精度等级
X3	1.0	0.031	好
X5	1.0	0.053	好
X8	1.0	0.148	好
X10	1.0	0.129	好
X11	1.0	0.167	好
X12	1.0	0.068	好

从表 2 可知, X3、X5、X8、X10、X11、X12 这 6 个特征等级评价都是“好”, 可以用于构建 SVR 算法预测模型。

我们将读取灰色预测得到的数据, 对数据进行标准化, 使用 SVR 算法预测所得预测值 1 和预测值 2 的建筑材料消耗量。如表 3 所示, 所得的数据与历史数据对比效果好说明建立的 SVR 模型拟合效果优良, 模型得到的预测值分别为 2,350,197 与 2,539,184, 并且模型的平均绝对误差与中值绝对误差相对较小, 可解释方差与 R2 值都十分接近 1, 这也表明模型的可信度较高。

Table 3. Predicting the results of numerical linear regression analyses
表 3. 预测数值线性回归分析结果

	线性回归分析结果(n = 15)						共线性诊断	
	非标准化系数		标准化系数	t	P	VIF	容忍度	
	B	标准误	Beta					
常数	-467.172	19102.416	-	-0.024	0.981	-	-	
预测值	0.997	0.015	0.999	68.257	0.000**	1.000	1.000	
R ²				0.997				
调整 R ²				0.997				
F	F (1,13) = 4659.034, p = 0.000							
D-W 值				1.746				

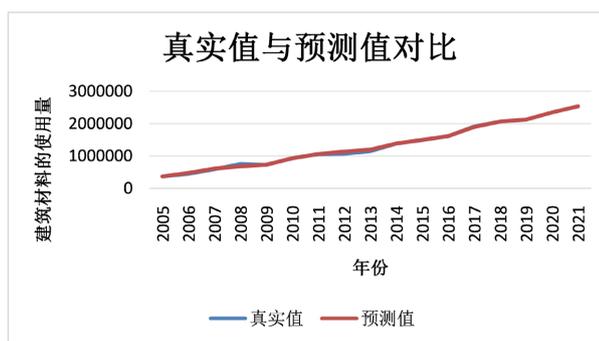


Figure 2. Plot of true vs. predicted values
图 2. 真实值与预测值的对比图

综上所述, 通过灰色预测算法与 SVR 算法来构建的建筑建材的预测体系与实际的发生数值高度重合如图 2, 表明了模型构建的正确性。灰色预测模型具有较高的预测精度和简单的模型检验方法, 但对数据光滑度要求较高; 而 SVR 算法能够处理非线性关系和高维问题, 并对异常点的偏差较小, 但计算复杂度较高。所以在数据的收集板块则要求数据的连续性与可读性, 在实际应用中, 可以根据数据特点和需求来选择合适的预测体系进行预测。基于预测数据, 企业可以调整建材采购计划, 以适应未来项目需求。通过优化采购策略, 企业可降低库存成本和资金占用成本, 提高资金利用效率。同时可优化生产资源配置和生产线布局。这有助于确保建材供应充足且及时, 提高生产效率和资源利用率, 以提升企业的经营效率和竞争力, 促进可持续发展。

5. 结论

建筑业作为国民经济的基石, 在社会发展中发挥着举足轻重的作用, 同时也消耗大量的建筑材料。推动建筑业高质量发展, 采取节能减排措施, 是实现我国“双碳”目标、节约大量经济成本和资源的必然要求。通过精细化管理优化建筑材料消耗是一项关键策略。这项工作的基础在于准确预测建筑材料需求, 从而支持优化施工作业管理和实现材料节约目标。

本文从建材的计算机算法预测的角度阐述了建筑建材影响因素的重要性及其精细化管理的必要性。在收集建筑历史数据的基础上, 采用数据相关性分析、使用 Lasso 提取特征、建立灰色预测模型和 SVR 模型, 最后通过 SVR 模型预测分析得出建筑建材的未来变化趋势, 进而对建筑企业的成本控制与节能降碳提出了建设性的措施, 为政府或企业进行宏观调控和各项政策的制定提供了一定的参考依据。

参考文献

- [1] 孙霞. 建筑节能与绿色施工管理的分析[J]. 科技创新与应用, 2013(4): 200.
- [2] 阮鹏. 建设工程绿色施工管理研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2015.
- [3] 赵小荣. 建筑施工管理与绿色建筑施工管理[J]. 江西建材, 2016(2): 285+290.
- [4] 钱仁兴. 基于绿色施工理念下建筑施工管理探析[J]. 科技创新与应用, 2017(17): 245.
- [5] 刘贵中. 绿色建筑施工管理在建筑施工管理中的应用探讨[J]. 科技创新与应用, 2019(31): 193-194.