

蔬菜类商品的自动定价与补货决策问题的研究

李宗友¹, 赵晓丽², 马涔绮³, 徐浩然¹, 李秉庚¹, 于昆平¹

¹沈阳航空航天大学航空宇航学院, 辽宁 沈阳

²沈阳航空航天大学理学院, 辽宁 沈阳

³沈阳航空航天大学人工智能学院, 辽宁 沈阳

收稿日期: 2023年10月3日; 录用日期: 2023年10月27日; 发布日期: 2023年11月6日

摘要

蔬菜类商品的销售与时间息息相关, 于是合理的销售组合对于蔬菜商品来说非常重要。本文为了得到合理的蔬菜商品的自动定价与补货模型, 根据一批蔬菜商品销售数据中各商品的历史销售和需求情况, 对接下来的蔬菜类商品销售情况进行预测和分析, 基于Apriori算法进行关联规则挖掘, 筛选出了符合蔬菜销售季节性要求、此时间段内可以销售的蔬菜商品, 接着用ARIMA模型进行时间序列预测。在此基础上通过需求价格弹性模型建立了定价模型, 得到了需求量变化和价格变化之间的关系。最后用粒子群寻优算法进行寻优求解, 寻找收益最大值, 得到了接下来一周在售的所有蔬菜的补货计划以及定价方案并且通过计算得到了最优方案收益。通过综合论证影响蔬菜商品的不同因素, 得出了竞争关系、用户反馈与历史需求等数据对商超蔬菜商品补货和定价决策的影响更大并分析了对其具体影响有哪些, 由此可以更加合理和精准的制定补货和定价策略。

关键词

Apriori算法, ARIMA模型, 粒子群寻优算法

Research on Automatic Pricing and Replenishment Decision-Making of Vegetable Commodities

Zongyou Li¹, Xiaoli Zhao², Cenqi Ma³, Haoran Xu¹, Bingeng Li¹, Kunping Yu¹

¹School of Aeronautics and Astronautics, Shenyang Aerospace University, Shenyang Liaoning

²School of Science, Shenyang Aerospace University, Shenyang Liaoning

³School of Artificial Intelligence, Shenyang Aerospace University, Shenyang Liaoning

Abstract

The sales of vegetable products are closely related to time, so the right sales mix is very important for vegetable products. In order to obtain a reasonable automatic pricing and replenishment model of vegetable commodities, this paper predicts and analyzes the sales of vegetable commodities according to the historical sales and demand of each commodity in a batch of vegetable commodity sales data, and performs correlation rule mining based on Apriori algorithm, screens out the vegetable commodities that meet the seasonal requirements of vegetable sales and can be sold within this time period, and then uses the ARIMA model to make time series forecasting. On this basis, the pricing model is established through the demand price elasticity model, and the relationship between demand change and price change is obtained. Finally, the particle swarm optimization algorithm is used to find the maximum return, and the replenishment plan and pricing scheme of all vegetables sold in the following week are obtained, and the optimal scheme benefit is obtained through calculation. Through the comprehensive demonstration of different factors affecting vegetable commodities, it is concluded that data such as competitive relationship, user feedback and historical demand have a greater impact on the replenishment and pricing decisions of vegetable commodities in supermarkets, and the specific impact on them is analyzed. This allows for more rational and accurate replenishment and pricing strategies.

Keywords

Apriori Algorithm, ARIMA Model, Particle Swarm Optimization Algorithm

Copyright © 2023 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

中国作为一个农业大国, 蔬菜的供给水平及其价格水平, 对农业经济效益和国民经济的健康和稳定产生重要影响。与此同时, 蔬菜价格水平以及稳定性对城乡居民的消费水平、生活质量产生重要影响。当然, 蔬菜作为一种商品不可能做到价格完全一成不变, 而是要达到一种动态的“定”, 按照市场经济理论所说, 即蔬菜的价格要围绕农产品自身的价值而上下波动。但最近几年农产品市场价格波动明显, 多种农产品的价格都在持续地暴涨, 其中以果蔬为代表的食用农产品市场价格波动最为明显。蔬菜价格也同样波动较大。价格巨幅波动增加了农户的收益风险和科学生产决策的难度。因此, 农产品价格尤其是蔬菜价格的波动逐渐成为社会关注的热点与焦点。

在一个生鲜超市中, 蔬菜类商品的特性包括短保鲜期和品质变化, 导致商品可能在一天内就无法再售。因此, 超市需要每天根据历史销售和需求情况进行补货和定价决策。然而, 蔬菜品种多样, 产地各异, 而进货交易通常在凌晨进行, 商家需要在不确切了解具体单品和进货价格的情况下, 制定当日蔬菜品类的补货计划。此外, 商超采用“成本加成定价”方法来定价商品, 对于品质变差或运损的商品通常会打折销售。为了做出明智的补货和定价决策, 商超需要可靠的市场需求分析。从需求角度看, 蔬菜类商品的销售量与时间存在关联关系; 从供给角度看, 蔬菜供应在4月至10月之间相对丰富, 但由于销售空

间的限制，确定合适的销售组合变得尤为关键。现有一批蔬菜商品销售数据，本文根据此数据研究分析。

2. 模型假设

- 1) 假设蔬菜的销售只受到主要因素的影响，即受到销售价格、季节、促销活动的影响，不受其他因素的干扰。
- 2) 假设蔬菜的销售与补货之间不存在延迟效应，即前一天的决策只对第二天的销售有影响。不考虑滞后效应对蔬菜销售的影响。
- 3) 假设市场需求是稳定的，在短期内不会发生剧烈变化，市场需求不受到其他因素的影响。

3. 蔬菜各品类相关性分析的研究

3.1. 数据预处理

对这批数据进行数据预处理，便于接下来的分析与建模，首先将退货的蔬菜销售数据剔除，其次数据聚合按照月份对销售数据进行分组，计算每个品类的总销售量和总销售额。最后，本文剔除数据中少于 20 周销量数据的蔬菜。表 1 展示预处理后的部分数据。

Table 1. Diagram of the practical teaching system of automation major

表 1. 各蔬菜品种销售量预处理结果

蔬菜品种	周数	1	2	7	8	9	10
	菜心		10.2	11.7	18.5	11.5	9.0
木耳菜		3.9	4.4	0.4	2.5	0.9	2.4
大白菜		12.7	18.2	17.8	16.5	14.6	14.3
云南油麦菜		22.3	41.4	27.9	23.7	19.8	10.3
青尖椒		3.7	2.8	4.0	4.4	3.2	2.6
红椒(1)		4.9	4.4	4.0	5.9	6.6	4.6
紫茄子(2)		18.5	20.7	18.9	13.1	9.0	7.6
青茄子(1)		18.6	18.5	8.4	7.6	8.7	8.5
西峡香菇(1)		14.5	13.4	12.5	11.9	11.1	11.4
金针菇(1)		15.4	13.7	12.5	12.3	12.3	10.2

3.2. 数据可视化

为了能更直观清晰的分析每个蔬菜品类和单品的销售量分布规律，使用直方图箱线图展示销售量的分布情况。

图 1 表述水生根茎类销售直方图的可视化，可以发现其中净藕(1)的销量最高，而野生粉藕、藕尖的销量为 0，则可以着重分析净藕(1)等销量额较高的蔬菜销量的关联性。

图 2 表述各个品种销售量箱型图，可以发现花叶类和辣椒类有几种蔬菜销量偏离均值较远，根据可视化可以着重分析此类销量。

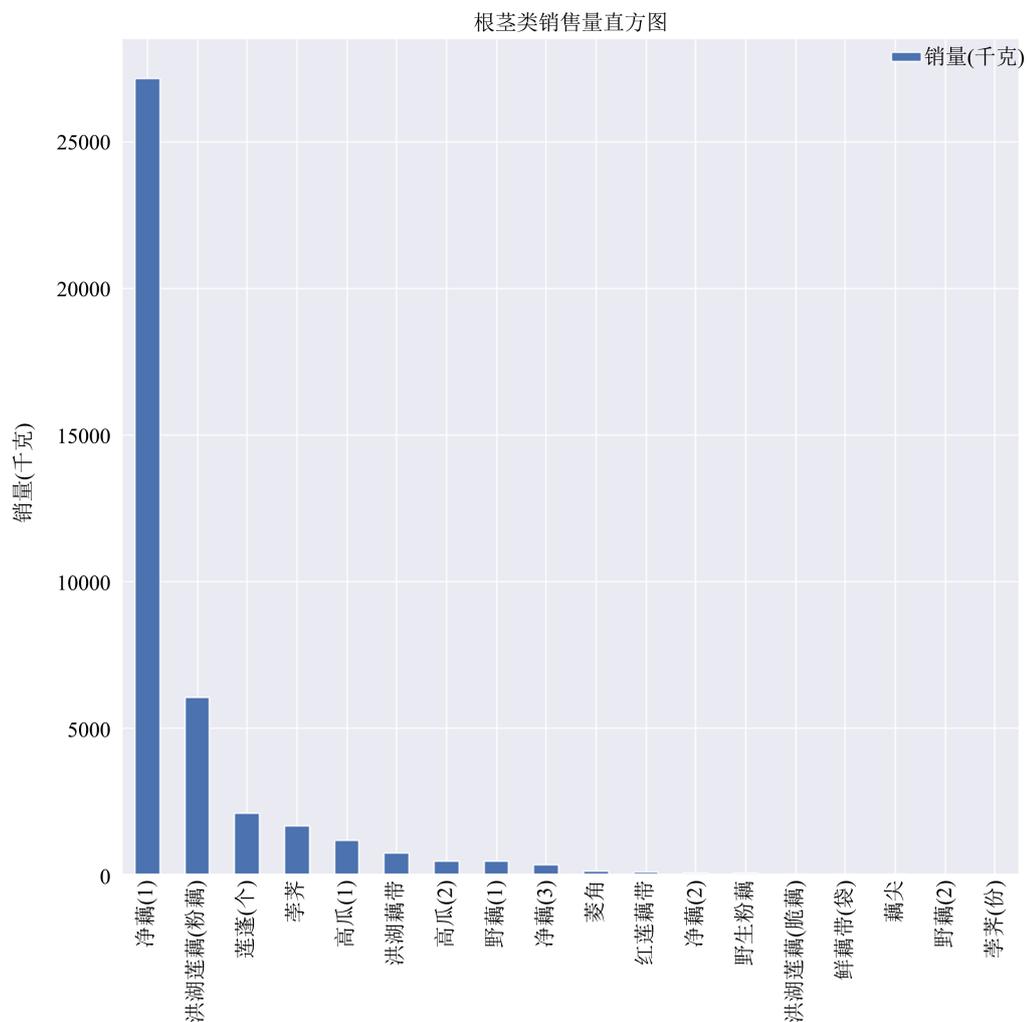


Figure 1. Diagram of the practical teaching system of automation major
图 1. 水生根茎类销售量直方图

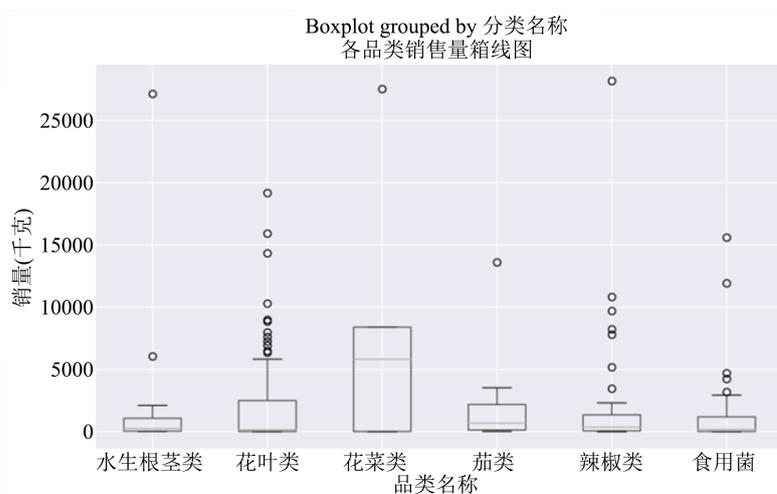


Figure 2. Diagram of the practical teaching system of automation major
图 2. 各品类销售量箱线图

3.3. 模型建立

本文问题需要找出蔬菜种类的相关性，基于此我们可以使用解决相关性问题的余弦相似度算法以及能寻找强关联规则的 Apriori 算法。

余弦相似度算法是指一个向量空间中两个向量夹角间的余弦值作为衡量两个个体之间差异的大小，余弦值接近 1，夹角趋于 0，表明两个向量越相似；余弦值接近于 0，夹角趋于 90 度，表明两个向量越不相似。

Apriori 算法[1]相对简洁明确，可以较好地发现数据之间的隐含规则，主要是通过循环迭代产生候选项集，在生成的候选集上经过剪枝操作产生频繁项集，进一步筛选符合条件规则集，最终将强关联规则输出。Apriori [2]算法的基本思想主要分为以下两个步骤：第一，查找频繁项集(支持度不低于最小支持度的项集为频繁项集)。第二，生成强关联规则。图 3 为 Apriori 算法流程图。

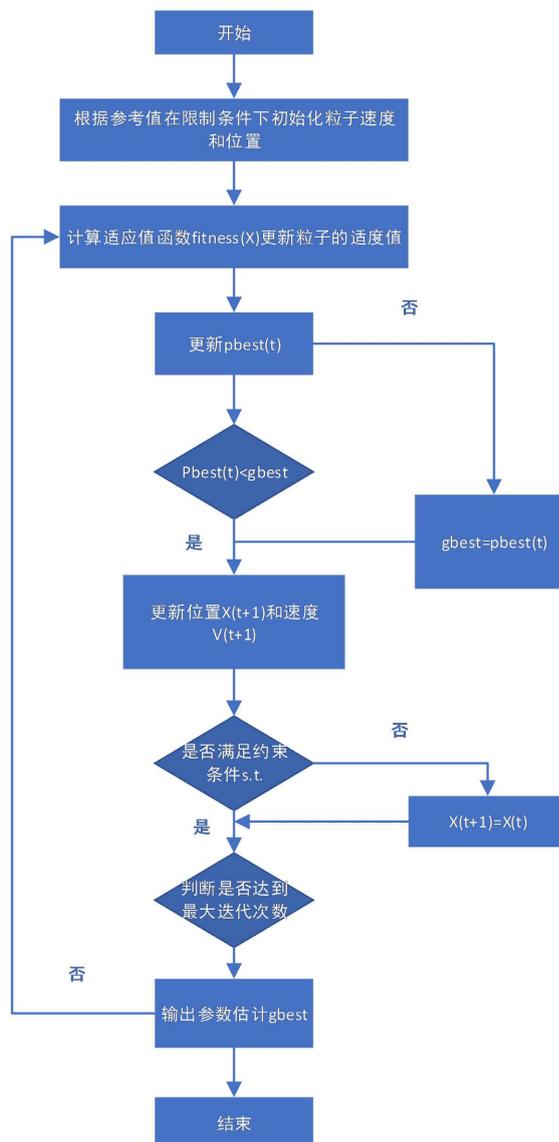


Figure 3. Apriori algorithm flowchart

图 3. Apriori 算法流程图

3.4. 模型求解

首先整理数据, 得到逐月品类的销量数据, 对数据进行整体分析, 得到大致销量以及在售的水平。选出主要销量数据的品种, 本文设置提取大于销售周期数大于 5 周数据菜品来做相关性分析, 表 2 是部分月份蔬菜品类的出售水平。

Table 2. Selling levels of each vegetable category in January and November

表 2. 1 月和 11 月各蔬菜品类出售水平

品类	1 月单品数	11 月单品数
花叶类	21	18
辣椒类	12	7
茄类	2	3
食用菌	10	3
花菜类	2	1
水生根茎类	3	2

在 MATLAB 中对数据进行余弦相似度算法计算, 分别对不同种类的蔬菜进行相关度分析。

分别遍历销售周期数大于五周的蔬菜种类和每个月份的销售总数据, 用 intersect 函数去求取两个向量或矩阵之间的交集。

选取 length(交集) > 1 时的数组, 对其进行余弦相似算法分析。公式实现如下:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (1)$$

其中 A , B 表示的是不同的蔬菜销量, $\cos(\theta)$ 的值反应了数据的相关程度, 值越接近 1 则表示其相关度越高。

将同月份中的 a , b 菜品销量带入公式, 得到了不同菜品之间的相关性。

对输出结果进行数据可视化, 使用 MATLAB 对数据进行热力图展示, 颜色最深的表示其相关性最高。如图 4、图 5 为部分数据展示。

对缺失值数据进行补充, 遍历每个单品, 每个月份取相关性在 0.9 以上的单品, 按照从大到小顺序依次对每个单品的数据进行关系式训练, 利用线性回归的方法补充缺失数据。

筛选数据对其进行 Apriori 算法分析, 按照月份筛选数据进行算法实现, 整理可得如表 3 所示部分蔬菜销量的关联规则。

按照相关性的大小和关联规则置信度的高低两个参数进行结合参考, 可以得到如表 4 所示蔬菜单品相关表。

结果可表示为, 在一月“云南生菜”“竹叶菜”“上海青”“牛首青菜”“茼蒿”“蔡甸藜蒿”的相关性最大, 顾客更倾向于同时买这些蔬菜种类。

对数据进行数据预处理可视化之后, 使用相关性分析以及使用关联规则挖掘算法对蔬菜的销售进行数据挖掘, 在实际进行销售的时候根据算法得出的关联性可以对销售策略以及销售方式进行调整。

捆绑销售作为一种重要的销售策略, 能帮助商家扩大销售, 增加利润。捆绑销售能产生多种协同作

用, 优势互补。产品之间的关联性越强, 所产生的协同作用就越大。比如使用的协同, 减少消费者的购买成本。

娃娃菜	0.48	0.55	0.63	0.75	0.78	0.82	0.43	0.4	0.43	0.7	0.59	0.76	0.93	0.7	0.53	0.75	0.95	0.73	0.64	0.78
枝江红菜苔	0.75	0.95	0.86	0.67	0.9	0.66	0.78	0.7	0.79	0.85	0.64	0.89	0.93	0.74	0.8	0.84	0.89	0.84	0.79	0.79
灯笼椒(1)	0.78	0.82	0.84	0.85	0.71	0.56	0.42	0.58	0.69	0.9	0.75	0.83	0.92	0.76	0.46	0.71	0.81	0.71	0.86	0.96
小米椒	0.85	0.7	0.87	0.73	0.8	0.46	0.71	0.7	0.75	0.96	0.94	0.97	0.92	0.84	0.52	0.89	0.71	0.87	0.95	0.85
螺丝椒	0.87	0.68	0.84	0.87	0.86	0.71	0.68	0.57	0.67	0.92	0.83	0.91	0.86	0.86	0.54	0.91	0.81	0.9	0.85	0.83
红灯笼椒(1)	0.79	0.78	0.87	0.73	0.7	0.56	0.69	0.82	0.78	0.7	0.57	0.81	0.8	0.84	0.84	0.67	0.76	0.76	0.73	0.85
姬菇(包)	0.86	0.97	0.94	0.69	0.72	0.69	0.84	0.99	0.95	0.83	0.3	0.74	0.65	0.76	0.99	0.66	0.9	0.74	0.87	0.57
东门口小白菜	0.84	0.99	0.92	0.6	0.87	0.7	0.9	0.93	0.9	0.91	0.91	0.94	0.98	0.71	0.93	0.8	0.89	0.83	0.92	0.72
银耳(朵)	0.64	0.65	0.76	0.7	0.72	0.65	0.65	0.49	0.52	0.84	0.76	0.85	0.95	0.71	0.68	0.73	0.81	0.74	0.91	0.7
奶白菜	0.69	0.76	0.81	0.84	0.86	0.76	0.71	0.55	0.61	0.91	0.59	0.84	0.95	0.78	0.64	0.83	0.97	0.84	0.82	0.66
荸荠	0.88	0.9	0.95	0.85	0.86	0.77	0.8	0.85	0.89	0.81	0.67	0.89	0.84	0.87	0.83	0.8	0.91	0.87	0.79	0.84
青梗散花	0.74	0.93	0.79	0.6	0.72	0.67	0.81	0.97	0.9	0.65	0.22	0.57	0.56	0.56	0.94	0.51	0.84	0.65	0.59	0.5
杏鲍菇(袋)	0.89	0.8	0.95	0.72	0.79	0.53	0.84	0.91	0.86	0.95	0.82	0.94	0.86	0.89	0.7	0.88	0.78	0.87	0.99	0.8
白玉菇(袋)	0.8	0.87	0.93	0.69	0.85	0.81	0.79	0.78	0.79	0.82	0.71	0.94	0.91	0.81	0.89	0.8	0.89	0.86	0.92	0.79
芜湖青椒(1)	0.92	1	0.98	0.93	0.92	0.56	0.89	1	0.89	0.98	0.86	0.95	0.88	0.97	0.59	0.98	0.88	0.97	0.92	0.74
洪湖莲藕(粉藕)	0.79	0.78	0.89	0.63	0.75	0.41	0.65	0.74	0.79	0.89	0.88	0.95	0.92	0.8	0.62	0.79	0.7	0.79	0.93	0.92
小青菜(1)	0.78	0.93	0.9	0.98	0.83	0.77	0.88	0.95	0.68	0.91	0.63	0.8	0.84	0.82	0.62	0.88	0.94	0.87	0.84	0.48
云南生菜(份)	0.47	0.75	0.66	0.97	0.76	0.96	0.35	0.97	0.41	0.74	0.66	0.77	0.9	0.68	0.67	0.66	0.98	0.66	0.59	0.91
云南油麦菜(份)	0.48	0.74	0.67	0.99	0.8	0.89	0.52	0.99	0.43	0.83	0.63	0.81	0.91	0.74	0.9	0.78	0.99	0.75	0.68	0.75
菠菜(份)	0.6	0.73	0.68	0.94	0.86	0.99	0.72	0	0.39	0.66	0.36	0.67	0.83	0.73	1	0.77	0.95	0.8	0.49	0.36
菜心(份)	0.81	0.73	0.74	0.94	0.81	0.99	0.41	0	0.6	0.69	0.57	0.7	0.84	0.72	0.28	0.66	0.97	0.7	0.48	0.8
上海青(份)	0.68	0.7	0.72	0.94	0.91	0.99	0.54	0	0.54	0.66	0.64	0.81	0.93	0.82	0.76	0.81	0.96	0.82	0.48	0.86
小米椒(份)	0.51	0.78	0.72	1	0.82	0.82	0.57	1	0.46	0.89	0.66	0.84	0.93	0.76	0.87	0.8	0.98	0.77	0.78	0.74
虫草花(份)	0.41	0.67	0.61	0.98	0.71	0.94	0.31	0.83	0.31	0.74	0.55	0.69	0.85	0.66	0.3	0.59	1	0.59	0.65	0.75
小皱皮(份)	0.46	0.74	0.63	0.99	0.78	0.9	0.62	0.78	0.34	0.75	0.49	0.68	0.86	0.69	0.4	0.69	1	0.7	0.71	0.48
青线椒	0.81	0.88	0.88	0.81	0.84	0.65	0.73	0.62	0.74	0.96	0.79	0.95	0.94	0.86	0.75	0.89	0.87	0.88	0.95	0.88
黄白菜(2)	0.95	0.93	0.99	0.84	0.88	0.78	0.95	0.99	0.97	0.89	0.57	0.92	0.95	0.84	0.91	0.81	0.93	0.9	0.87	0.89
金针菇(盒)	0.54	0.85	0.76	0.95	0.83	0.66	0.48	1	0.52	0.95	0.75	0.87	0.97	0.82	0.71	0.81	0.91	0.76	0.92	0.89
海鲜菇(袋)(4)	0.94	0.98	0.98	0.81	0.89	0.9	0.93	0.98	0.97	0.87	0.93	0.97	0.96	0.82	0.99	0.81	0.97	0.89	0.89	0.83

Figure 4. Correlation analysis of sales volume of major dishes in January

图 4. 1 月主要菜品销售量相关性分析

西兰花	0.87	0.91	0.98	0.94	0.98	0.96	0.94	0.71	0.87	0.88	0.87	0.94	0.91	0.75	1	0.83	0.89	0.91		
黄心菜(1)	0.81	0.72	0.85	0.81	0.82	0.86	0.94	0.68	0.74	0.89	0.78	0.68	0.85	0.72	0.83	1	0.8	0.72		
平菇	0.94	0.67	0.93	0.98	0.95	0.94	0.85	0.97	0.93	0.93	0.97	0.92	0.95	0.52	0.89	0.8	1	0.82		
净藕(1)	0.83	0.87	0.89	0.82	0.93	0.9	0.81	0.56	0.73	0.78	0.77	0.94	0.74	0.88	0.91	0.72	0.82	1		
红杭椒	0.89	0.83	0.95	0.96	0.95	0.97	0.9	0.93	0.93	0.9	0.93	0.88	0.98	0.62	0.93	0.89	0.91	0.76		
菠菜	0.93	0.84	0.94	0.96	0.9	0.97	0.88	0.87	0.9	0.88	0.94	0.75	0.95	0.56	0.88	0.74	0.93	0.79		
娃娃菜	0.21	0.82	0.73	0.99	0.62	0.96	0.59	0.15	0.23	0.24	0.21	0.82	0.29	0.77	0.6	0.34	0.98	0.78		
枝江红菜苔	0.83	0.78	0.91	0.88	0.9	0.94	0.8	0.67	0.76	0.87	0.83	0.9	0.84	0.66	0.92	0.71	0.89	0.9		
小米椒	0.97	0.83	0.99	0.96	0.97	0.97	0.97	0.84	0.94	0.98	0.97	0.97	0.96	0.65	0.98	0.88	0.94	0.88		
螺丝椒	0.98	0.83	0.96	0.92	0.9	0.94	0.9	0.8	0.92	0.94	0.96	0.78	0.89	0.73	0.88	0.8	0.91	0.86		
随州泡泡青	0.4	0.89	0.62	0.7	0.85	0.92	0.54	0.13	0.44	0.52	0.42	0.95	0.56	0.77	0.87	0.77	0.69	0.85		
东门口小白菜	0.89	0.79	0.77	0.72	0.8	0.79	0.65	0.38	0.72	0.77	0.81	0.88	0.61	0.81	0.73	0.41	0.76	0.96		
奶白菜	0.91	0.84	0.93	0.91	0.87	0.96	0.87	0.73	0.79	0.95	0.91	0.81	0.89	0.74	0.89	0.92	0.9	0.85		
白玉菇(袋)	0.87	0.56	0.82	0.93	0.83	0.86	0.7	0.98	0.91	0.84	0.93	0.67	0.88	0.59	0.74	0.7	0.97	0.72		
芜湖青椒(1)	0.49	0.86	0.86	0.98	0.94	0.98	0.72	0.95	0.66	0.53	0.51	0.98	0.78	0.77	0.95	0.94	0.9	0.87		
洪湖莲藕(粉藕)	0.27	0.65	0.58	0.98	0.84	0.85	0.29	1	0.48	0.46	0.24	0.94	0.77	0.6	0.86	0.97	1	0.75		
小青菜(1)	0.99	0.92	0.95	0.94	0.76	0.97	0.97	0.35	0.99	0.83	0.94	0.61	0.82	0.85	0.77	0.41	0.82	0.84		
圆茄子(2)	0.87	0.66	0.91	0.94	0.91	0.92	0.89	0.9	0.89	0.92	0.91	0.83	0.95	0.51	0.87	0.94	0.92	0.71		
青线椒	0.96	0.79	0.98	0.98	0.97	0.97	0.95	0.88	0.97	0.97	0.98	0.95	0.97	0.64	0.96	0.88	0.95	0.84		
黄白菜(2)	0.95	0.73	0.93	0.97	0.96	0.92	0.81	0.95	0.96	0.88	0.98	0.94	0.93	0.63	0.9	0.89	0.96	0.84		

Figure 5. Correlation analysis of sales volume of major dishes in November

图 5. 11 月主要菜品销售量相关性分析

Table 3. Some vegetable correlation displays**表 3.** 部分蔬菜关联性展示

关联规则	置信度
小米椒--->枝江青梗散花	0.6
小米椒--->金针菇	0.79
西峡香菇--->芜湖青椒	0.6
芜湖青椒--->紫茄子	0.36

Table 4. Table related to vegetable items**表 4.** 蔬菜单品相关表

一月	二月	三月	四月	五月	六月	七月
云南生菜	白菜苔	四川红香椿	四川红香椿	牛首生菜	西峡花菇(1)	苋菜
竹叶菜	苋菜	西峡花菇(1)	西峡花菇(1)	西峡花菇(1)	苋菜	云南生菜
上海青	云南生菜	白菜苔	苋菜	苋菜	云南生菜	竹叶菜
牛首油菜	竹叶菜	苋菜	云南生菜	云南生菜	竹叶菜	小白菜
茼蒿	上海青	云南生菜	竹叶菜	竹叶菜	小白菜	上海青
蔡甸藜蒿	萝卜叶	竹叶菜	小白菜	小白菜	上海青	菜心
菜心	牛首油菜	上海青	上海青	上海青	菜心	木耳菜
大白菜	茼蒿	萝卜叶	萝卜叶	菜心	木耳菜	大白菜
云南油麦菜	蔡甸藜蒿	茼蒿	茼蒿	木耳菜	云南油麦菜	云南油麦菜
青尖椒	菜心	蔡甸藜蒿	蔡甸藜蒿	云南油麦菜	红尖椒	红尖椒
红椒(1)	大白菜	菜心	菜心	红尖椒	青尖椒	青尖椒

在蔬菜市场上的捆绑销售大致方法主要采取同质商品捆绑销售、互补式商品捆绑销售。同质商品捆绑销售这种方式一般是将若干数量的同质商品集合包装出售，互补式商品捆绑销售即把消费上具有关联性、用途上具有互补性的商品组合起来，成为一单位商品包销售给消费者。

例如一月将“云南生菜”“竹叶菜”“上海青”“牛首青菜”“茼蒿”等蔬菜，二月将“白菜苔”“苋菜”“云南生菜”“竹叶菜”“上海青”等蔬菜在实际销售时进行售卖位置临近处理，在六月对“云南油麦菜”“木耳菜”“红尖椒”进行蔬菜捆绑销售等提高销量。

4. 蔬菜补货与定价决策模型构建

4.1. 基于 ARIMA 模型的蔬菜预测模型

4.1.1. 数据预处理

通过对数据的清洗和数据筛选，剔除异常数据和在销售中被退货的数据，假如我们要求解在 2023 年 7 月 1 日~7 日的蔬菜预测模型，我们首先选出其中在 2023 年 7 月 1 日~7 日之前的五周均售卖的蔬菜，避免季节性因素对预测结果的干扰。从 2020 年 7 月 1 日开始，按照以每周作为最小单位统计其平均销量、平均价格等信息并对其进行预测。

4.1.2. ARIMA 模型

ARIMA 模型[3] [4] (Autoregressive Integrated Moving Average model)，即差分整合移动平均自回归模

型, 又称整合移动平均自回归模型, 是时间序列预测分析方法之一。ARIMA 是在平稳的时间序列基础上建立起来的, 因此时间序列的平稳性是建模的重要前提。如果时间序列不稳定, 也可以通过一些操作去使得时间序列稳定(比如取对数, 差分), 然后进行 ARIMA 模型预测, 得到稳定的时间序列的预测结果, 最后对预测结果进行之前使序列稳定的操作的逆操作(取指数差分的逆操作)就可以得到原始数据的预测结果。其具体过程如下:

将自回归模型(AR)、移动平均模型(MA)和差分法结合, 我们就得到了差分自回归移动平均模型 ARIMA。

用公式表示为:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} \quad (2)$$

在本题中, 我们可以通过 MATLAB 软件进行计算, 由此可以得到每个种类的蔬菜对应的 ARIMA 模型的结构参数, 进而计算其模型的公式。例如, 我们在预测西峡花菇(1)在 2023 年 7 月 1 日~7 日的平均销量、平均售价、平均批发价格时, 可以分别得到其对应的模型:

$$X_t = -0.8418 - 0.5177X_{t-1} - 0.0199X_{t-2} + 1 * \varepsilon_{t-1} + \varepsilon_t \quad (3)$$

$$X_t = -5.6514 - 0.7825X_{t-1} + 1 * \varepsilon_{t-1} + \varepsilon_t \quad (4)$$

$$X_t = 0.0047 + 0.9177X_{t-1} - 0.1227X_{t-2} + 1 * \varepsilon_{t-1} + \varepsilon_t \quad (5)$$

通过以上求解计算可以得到 36 种蔬菜在 2023 年 7 月 1 日~7 日的平均销售量、平均价格、平均批发价格预测值。如图 6、图 7、图 8 所示。

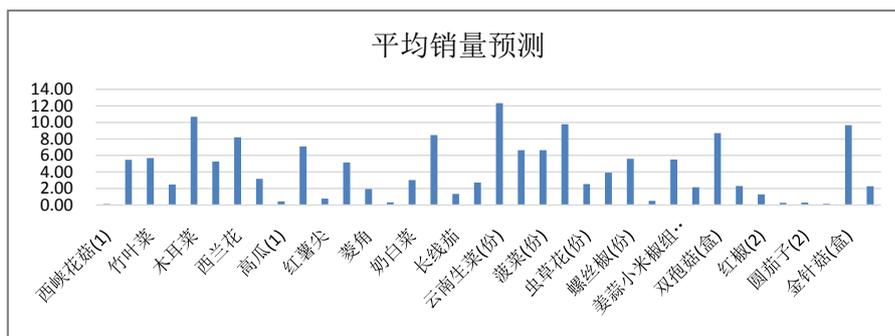


Figure 6. Average sales forecast
图 6. 平均销量预测

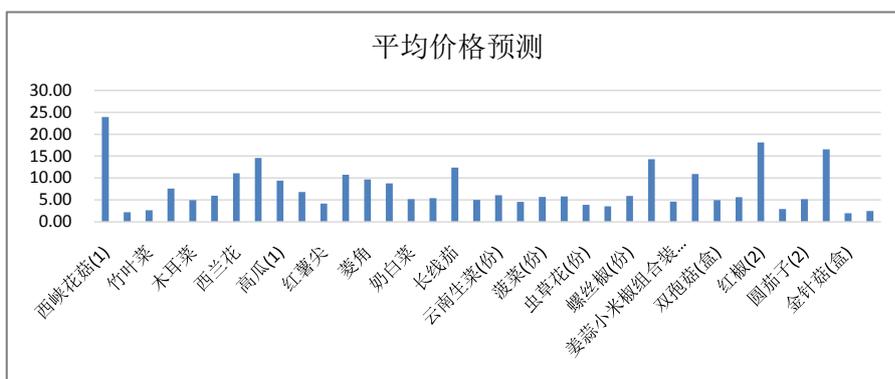


Figure 7. Average price forecast
图 7. 平均价格预测



Figure 8. Average wholesale price forecast
图 8. 平均批发价格预测

4.2. 需求价格弹性模型的构建

需求价格弹性[5] [6]是表示在一定时期内一种商品的需求量变动对于该商品的价格变动的反应程度。我们通常用需求价格弹性系数加以表示，即：

$$\text{需求价格弹性系数} = \frac{\text{价格变动的百分比}}{\text{需求量变动的百分比}}$$

以 E 表示需求价格弹性系数， Q 和 ΔQ 分别表示商品的需求量和需求的变动量， P 和 ΔP 分别表示价格和价格的变动量，则需求价格弹性系数的一般表达式为：

$$E_t = \frac{\Delta P}{\Delta Q} = \frac{P_t - P_{t-1}}{Q_t - Q_{t-1}} \quad (6)$$

在本问中，我们使用 MATLAB 软件进行计算，得到各品种蔬菜的需求价格弹性模型的 E 值，如图 9 所示。假设自变量为进货量，除去损耗之外就是销量。此时需求价格弹性系数已知，销量变化多少，价格也会变化多少，反之同理，由此可以得到销量变化和价格变化的关系。

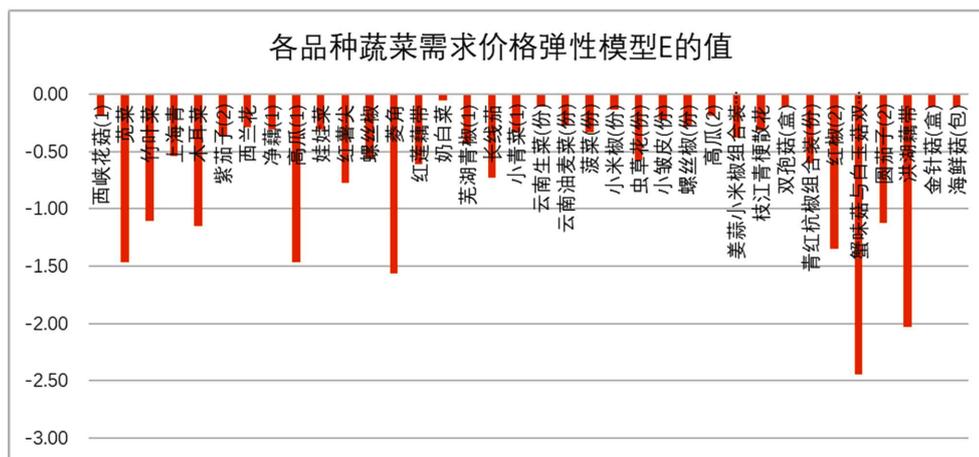


Figure 9. The value of the price elasticity model E of demand for each variety of vegetables
图 9. 各品种蔬菜需求价格弹性模型 E 的值

4.3. 粒子群算法——函数寻优

4.3.1. 模型建立

粒子群优化算法[7]的优化策略为根据现在所求得的局部最优解和全局最优解进行迭代，使得整个粒

子群向着最优解的方向移动,使得整体的适应值(性能指标)不断的优化,直到达到我们所设定的标准。再提取最后优化过的全局最优解作为我们所求得的结果,这样就实现了求取最优解的目标。

粒子可以根据局部最优解和全局最优解来更新自己的速度和位置。

$$V_i(t+1) = \omega V_i(t) + c_1 * rand() * (pbest_i - X_i(t)) + c_2 * rand() * (gbest - X_i(t)) \quad (7)$$

$$X_i(t+1) = X_i(t) + V_i(t) \quad (8)$$

式中, $i=1,2,\dots,N$, N 是此群粒子的总数为 200; t 为迭代次数, $t = 500$; $rand()$ 是介于(0,1)之间的随机数; c_1 和 c_2 是学习因子, $c_1 = 0.4$, $c_2 = 0.6$, X_i 和 V_i 分别表示 i 粒子的速度和位置。

算法步骤如图 10 所示:

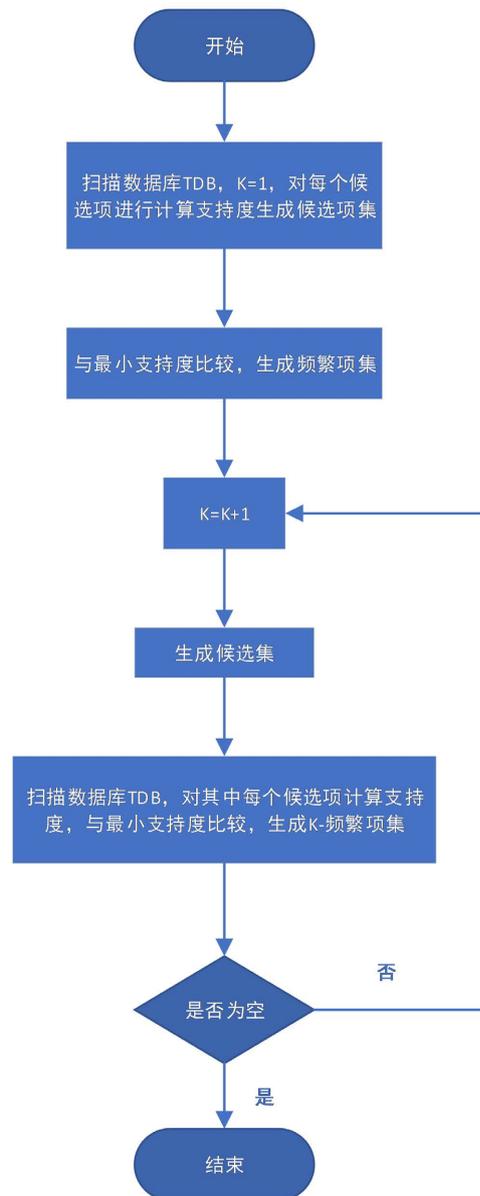


Figure 10. Particle swarm optimization algorithm flowchart

图 10. 粒子群优化算法流程图

本文对于该蔬菜定价及补货策略多问题寻优的目标函数为：

Step1. 计算实际销量

$$M = J_0 * (1 - H) \quad (9)$$

Step2. 计算需求变化量

$$L = M - X_0 \quad (10)$$

Step3. 计算销售价格

$$N = S_0 + L * E \quad (11)$$

Step4. 得到目标函数

$$f(x) = \sum_{i=1}^n (M * N - J * P) * 7 \quad (12)$$

Step5. 制定约束条件

$$27 < a < 33 \quad (13)$$

$$\max_x f(x) \quad x \in a \quad (14)$$

其中 H : 损耗率; X_0 : 销售量初始值; J_0 : 进货量初始值; S_0 : 销售价格初始值; P : 批发价格; a : 蔬菜种类。

4.3.2. 模型求解

在本文中，我们通过粒子群算法寻优进货量，预测得到的 2023 年 7 月 1 日~7 日的销量作为初始的需求量，设置预测的销售价格，批发价格同理，设置进货量范围为需求量的 0.5~1.5 倍，进货量作为变量，要想得到需求量，需要减掉消耗的，需求量相对初始需求量变化多少，计算出新的销售价格，减去批发成本得到收益，通过粒子群算法寻找蔬菜收益最大值，寻优结果如图 11 所示：

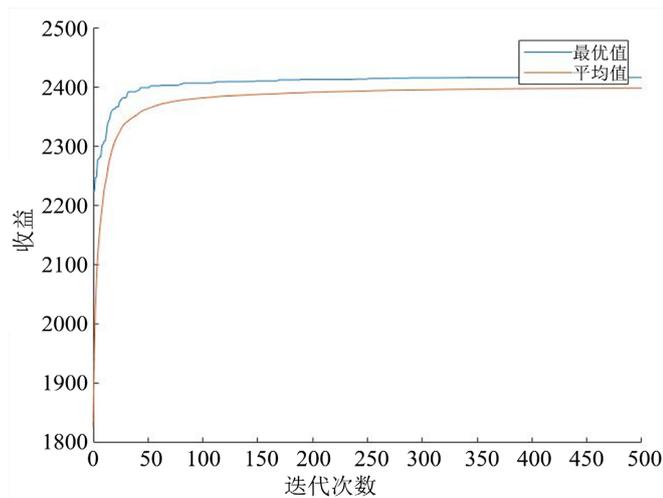


Figure 11. Particle swarm optimization results

图 11. 粒子群算法寻优结果

随着迭代次数的增多，收益值趋于稳定，大约在 2400 元左右。求解模型可得 7 月 1 日~7 日在售的所有蔬菜的补货计划以及定价，计算可得最优方案收益如表 5 所示，最优方案收益为：2411.8 元。

Table 5. Purchase planning and pricing strategy
表 5. 进货计划及定价策略

菜品	名称	进货量	损毁量	销量	销售价格	收入	成本	利润
食用菌	西峡花菇(1)	0.19	0.02	0.17	23.91	3.96	2.90	1.07
花叶类	苋菜	3.33	0.62	2.71	6.24	16.92	5.10	11.82
花叶类	竹叶菜	3.56	0.48	3.07	5.50	16.93	6.18	10.75
花叶类	上海青	3.70	0.53	3.17	7.22	22.87	14.05	8.82
花叶类	木耳菜	7.10	0.54	6.56	9.72	63.72	13.33	50.39
茄类	紫茄子(2)	5.76	0.35	5.41	5.90	31.90	18.86	13.04
花菜类	西兰花	9.62	0.89	8.73	10.95	95.57	71.92	23.65
水生根茎类	净藕(1)	4.77	0.26	4.51	14.18	63.88	49.21	14.67
水生根茎类	高瓜(1)	0.22	0.07	0.16	9.84	1.56	2.50	-0.94
花叶类	娃娃菜	8.13	0.20	7.93	6.56	52.04	36.17	15.87
花叶类	红薯尖	1.18	0.10	1.08	3.97	4.29	1.58	2.71
辣椒类	螺丝椒	3.01	0.31	2.70	12.16	32.85	29.64	3.21
水生根茎类	菱角	1.42	0.14	1.29	10.74	13.81	10.40	3.41
水生根茎类	红莲藕带	0.50	0.08	0.42	8.71	3.63	2.40	1.23
花叶类	奶白菜	4.39	0.69	3.70	5.14	19.04	11.21	7.84
辣椒类	芜湖青椒(1)	6.14	0.35	5.79	6.57	38.06	23.78	14.28
茄类	长线茄	2.00	0.14	1.87	12.00	22.40	13.97	8.43
花叶类	小青菜(1)	4.07	0.42	3.65	4.67	17.05	10.86	6.19
花叶类	云南生菜(份)	16.39	1.55	14.84	5.84	86.67	56.45	30.22
花叶类	云南油麦菜(份)	6.79	0.64	6.15	4.70	28.89	19.40	9.49
花叶类	菠菜(份)	6.04	0.57	5.47	6.06	33.16	22.81	10.35
辣椒类	小米椒(份)	14.68	1.38	13.30	5.30	70.42	31.00	39.42
食用菌	虫草花(份)	2.80	0.26	2.54	3.87	9.81	7.19	2.62
辣椒类	小皱皮(份)	4.49	0.42	4.07	3.50	14.24	10.52	3.72
辣椒类	螺丝椒(份)	4.96	0.47	4.49	6.25	28.06	22.95	5.11
水生根茎类	高瓜(2)	0.26	0.02	0.24	14.33	3.43	3.96	-0.53
辣椒类	姜蒜小米椒组合装(小份)	5.98	0.56	5.41	4.66	25.23	13.88	11.35
花菜类	枝江青梗散花	3.22	0.30	2.92	10.68	31.16	23.87	7.28
食用菌	双孢菇(盒)	9.91	0.02	9.89	4.83	47.81	34.10	13.71
辣椒类	青红杭椒组合装(份)	2.65	0.25	2.40	5.55	13.34	9.50	3.84
辣椒类	红椒(2)	1.83	0.17	1.66	17.59	29.18	24.19	4.99
食用菌	蟹味菇与白玉菇双拼(盒)	0.13	0.00	0.13	3.28	0.44	0.35	0.08

Continued

茄类	圆茄子(2)	0.44	0.03	0.41	5.07	2.08	1.58	0.50
水生根茎类	洪湖藕带	0.08	0.02	0.06	16.72	0.97	1.27	-0.29
食用菌	金针菇(盒)	8.55	0.04	8.51	2.06	17.54	12.37	5.16
食用菌	海鲜菇(包)	2.80	0.00	2.80	2.41	6.76	5.69	1.07

5. 对其他影响蔬菜销售的因素的分析

影响商超补货和定价的原因有多种[8], 补货量与价格主要是由于消费者需求量的变化而改变, 而多方面因素也导致着用户需求量发生变化, 一般说来蔬菜需求量的影响因素主要包括定量因素和定性因素。定性因素是不容易被量化的因素, 例如宏观政策、经济发展趋势、城市发展水平等; 定量因素是可量化的因素, 主要有季节性、天气、蔬菜价格、节假日、替代品价格、总人口数、恩格尔系数、人均可支配收入等。在对蔬菜需求量的预测中, 短期内总人口数、恩格尔系数、人均可支配收入等定量因素比较稳定, 所以结合相关研究中的主要影响因素、蔬菜本身的特点和供需特点, 需要考虑的需求影响因素主要包括有天气、节假日、个人偏好、历史需求等等。我们主要从竞争关系、用户反馈、历史需求三个方面进行主要阐述。

5.1. 竞争关系

竞争关系方面。需求是蔬菜生产的前提, 随着蔬菜市场的不断发展, 竞争更加激烈, 同时随着消费者消费观念和消费能够的转变, 不断适应消费者的需求, 提高消费者的忠诚度和满意度成为蔬菜生产主体竞争的重要手段。商超需要时刻关注同行价格, 查看同行有无同类替代菜品等价格数据。

5.2. 用户反馈

顾客反馈与服务体系。顾客反馈数据在产业服务体系中发挥巨大作用, 是顺利实现产品生产、加工、流通、交易的基本前提, 对产业链的每一部分都影响巨大。这种作用可能直接满足或增加消费者需求, 可能直接减少生产成本或交易成本, 可能改善产业生产、加工活动主体的效率, 可能减少了组织管理成本, 可能是作用于产业环境而减少产业整体生产加工成本或效率。

5.3. 历史需求

要结合当地环境与历史需求。生产布局对生产效益的影响深远, 当地属于什么气候类型, 人们的喜爱偏好都是造成蔬菜需求量变化的重要因素。根据比较优势理论和资源禀赋理论, 同时结合当地喜好特点、习俗进行适当进货与定价。资源与蔬菜品种的生长相适应, 或者生产区域具有运输、加工、消费的区位优势, 都可能对蔬菜产量、生产成本、运输成本以及交易成本产生影响, 进而引起生产效益的变化。

通过实现蔬菜需求预测系统, 对具有短生命周期特性的蔬菜需求量进行准确把控, 从而为制定蔬菜的生产流通规划提供重要依据。掌握蔬菜需求量的变化规律, 了解蔬菜需求的趋势, 将有利于正确引导蔬菜流通和物流运输计划, 实现蔬菜市场的供需平衡, 有效解决买者与卖者之间的信息不对称问题, 有效提高农民效益和维护城市居民切身利益。

6. 结论

本文基于现有的一批蔬菜商品销售数据制定了一套合适的销售组合, 以 2023 年 7 月 1 日~7 月 7 日为例, 首先通过余弦相似度算法以及 Apriori 算法找出蔬菜种类的相关性, 建立基于 ARIMA 模型的蔬菜

预测模型, 求解计算得到 36 种蔬菜在 2023 年 7 月 1 日~7 日的平均销售量、平均价格、平均批发价的预测值。其次建立需求价格弹性模型, 得到销量变化和价格变化的关系。最后通过粒子群算法求得 2023 年 7 月 1 日~7 日蔬菜收益最大值。应用蔬菜自动定价与补货方案, 可以有效解决生鲜商超中保鲜期短且品相随销售时间的增加而变差等蔬果类商品当日未售出隔日就无法再售这一问题, 为蔬果供应链的良好发展提供可行性思路, 在保证商超获得利益的同时, 也减少了资源浪费, 以便更好的实行可持续发展。这些模型相互之间配合紧密, 所得结果依次递进, 使最终求解真实可靠, 模型充分联系实际, 具有很好的通用性和推广性。

基金项目

辽宁省教育厅青年科技人才“育苗”项目[JYT2020128]; 2022 年度辽宁省普通高等教育本科教学改革研究优质教学资源建设与共享项目: 跨校修读学分背景下《高等数学》教学模式研究。

参考文献

- [1] 高宾, 王兰成. 基于 Apriori 数据挖掘算法的网络舆情信息索引研究[C]//中国索引学会. 2019 年中国索引学会年会暨学术研讨会论文集: 2019 年卷. 上海: 复旦大学图书馆, 2019: 72-77.
- [2] 罗章铭. 基于二进制编码与增量更新的 Apriori 算法优化研究[D]: [硕士学位论文]. 长沙: 湖南师范大学, 2022.
- [3] 夏如玉, 王梓桥. 基于 ARIMA 模型对重庆市 GDP 预测分析[J]. 中国储运, 2022(8): 93-94.
- [4] 刘峰, 王儒敬, 李传席. ARIMA 模型在农产品价格预测中的应用[J]. 计算机工程与应用, 2009, 45(25): 238-239, 248.
- [5] 蒋志芳. 需求价格弹性理论的一个应用——大学生手机消费的实证分析[J]. 江苏商论, 2012(6): 7-10.
- [6] 王雪青, 孙妩. 需求弹性理论分析及企业决策应用[J]. 技术经济与管理研究, 2008(3): 38-40.
- [7] 陈建新. 基于改进粒子群算法的函数优化研究[D]: [硕士学位论文]. 长春: 长春理工大学, 2019.
- [8] 毛莉莎. 供应链视角下蔬菜批发市场定价策略及产销模式研究[D]: [硕士学位论文]. 长沙: 中南林业科技大学, 2022.