蔬菜类商品的自动定价与补货决策研究

芦星宇 蒙雨 董明 史可鉴 朱轩平
航天工程大学

摘 要： 如何建立蔬菜类商品定价与补货模型，厘清蔬菜销售量、供应品种和定价补货之间的关系，从而解决海量蔬菜市场存在的保鲜期短的实际问题，具有重要意义。结合实际情况对蔬菜类商品销售情况做出合理假设，利用 Apriori 算法建立蔬菜品类的关联模型，针对品类和单品不同情形，利用线性回归、ARIMA、非线性规划方法以及梯度提升决策树算法建立蔬菜定价的决策模型，可以给出不同情形下使商超收益最大的补货与定价策略。

关键词： KS 检验；Apriori 算法；ARIMA 模型；非线性规划；遗传算法

【DOI】10.12229/j.issn.1672-5719.2024.05.010

作者简介： 芦星宇（2002-），男，汉族，内蒙古乌兰浩特人，本科在读，航天测控技术（雷达工程）专业；蒙雨（2002-），男，汉族，吉林松原人，本科在读，航天测控技术（雷达工程）专业；董明（2000-），男，汉族，黑龙江大庆人，本科在读，航天测控技术（雷达工程）专业；史可鉴（2002-），男，汉族，四川绵阳人，本科在读，航天测控技术（雷达工程）专业；朱轩平（2000-），男，汉族，江苏常州人，本科在读，航天测控技术（雷达工程）专业。

一、 问题重述

（一）问题背景

在生鲜商超中，蔬菜类商品的保鲜期较短，品质随销售时间增加而下降。通常情况下，商超每天根据历史销售和需求情况进行蔬菜补货。由于蔬菜品种繁多且产地各异，商家需要在凌晨 3:00—4:00 进行进货交易，但并不清楚具体单品和进货价格。商超采用成本加成定价方法对蔬菜进行定价，对于有运损和品质下降的商品通常会打折销售。准确的市场需求分析对补货和定价决策非常重要。从需求角度来看，蔬菜销售量往往与时间存在一定的关联；从供应角度来看，4 月至 10 月是蔬菜供应品种较多的季节。由于商超销售空间有限，合理的销售组合变得尤为重要

（二）问题重述

问题一：基于所提供的数据，分析蔬菜类商品不同品类或不同单品之间可能存在的关联关系，及蔬菜各品类和单品销售量的分布规律。

问题二：商超以品类为单位进行补货计划，需要分析各蔬菜品类的销售总量与成本加成定价之间的关系。然后，根据分析结果，给出未来一周（2023 年 7 月 1 日至 7 日）各蔬菜品类的日补货总量和定价策略，以最大化商超的收益。

问题三：商超希望更精确的制定单品的补货计划，要求可售单品总数控制在 27-33 个，并且各单品的订购量必须满足最小陈列量 2.5 千克的要求。根据 2023 年 6 月 24 日至 6 月 30 日的可售品种，我们需要给出 7 月 1 日的单品补货量和定价策略，以在尽量满足市场对各品类蔬菜商品需求的前提下，实现商超最大化的收益。

问题四：商超还需要采集哪些相关数据，以更好地制定蔬菜商品的补货和定价决策。

二、基本假设

假设 1：同一天同一时刻认定为同一订单，且若十秒内没有新的购买，认为进入下一个订单；不考虑商超中多个收银台同时收银的情况。

假设 2：假设销售量、价格和损耗率在短期内保持相对稳定，不会受到外部因素（如季节性变化、市场突发事件等）的剧烈波动。

假设 3：忽略市场上其他竞争对手的影响以及同品类蔬菜的替代效应。

假设 4: 假设在所给的条件下, 存在一个唯一最优解, 即可以得到一个最佳的补货定价策略，以最大化商超效

益。

假设 5：假设商超的上级仓储机构货源充足，不会出现货物短缺的情况。

三、符号说明

符号说明（见表 1）。

表1 符号说明

符号	说明
cost_j	表示 j 种蔬菜的成本
$sale_{i,j}$	表示第 i 天，j 种蔬菜的销售价格
$sale\_vol_{i,j}$	表示第 i 天，j 种蔬菜的销量
$whole\_price_{i,j}$	表示第 i 天，j 种蔬菜的批发价格
$wastage_j$	表示 j 种蔬菜的损耗率
$\omega_{h,d}$	表示商 品 h 在超市 d 的价格
$\omega_{h,r}$	表示商品 h 在竞争者 r 的价格
$\tau_{h,r}$	表示各个竞争者 价格的权重
$\delta_{d,r}$	表示竞争者 r 的价格印象与超市 d 价格印象的差
$q_{h,r}$	表示商品 h 在竞争者 r 的销售量
$q_{h,d}$	表示商品 h 在超市 d 的销售量

四、问题的分析

（一）问题一的分析

解决问题一，要分析商品不同品类和不同单品间的关联关系。为了减小误差，我们将每时刻的流水划分为一个订单，运用 Apriori 算法将不同订单之间的数据进行关联分析，可以避免不同日期间的购买情况对分析产生干扰的问题。而后运用统计学原理，将各品类及单品销售量的规律运用图像的方式直观表达。

（二）问题二的分析

为解决问题二，我们将采用成本加成定价的方法来确定定价结果。首先，我们会根据不同品类的成本和预期利润率，计算出每个品类的建议销售价格。然后，我们将使用对数函数、幂函数、指数函数、线性函数等来拟合销售总量与平均销售定价之间的关系，以便分析数据并计算出每天的平均批发价。接下来，我们将使用 ARIMA 模型对未来一周的销售情况进行预测。ARIMA 模型是一种时间序列模型，可以根据过去的销售流水数据预测未来的销售趋势和变化。通过使用 ARIMA 模型，我们可以获得下一周的销售相关信息，包括销售量和销售价。最后，我们将使用非线性优化模型制定定价策略。该模型将考虑多个因素，如成本、利润、市场需求等，以最大化利润为目标。通过对定价策略进行优化，我们可以找到最佳的定价方案，以实现最大化利润的目标。

（三）问题三的分析

为解决问题三，我们将在问题二的基础上增加题中给出的约束条件。首先，我们将对上一周的销售情况进行分析与统计，以了解销售量、销售额等相关指标的情况。通过对销售数据的分析，我们可以获取关键的统计信息，如平均销售量、销售额的波动情况、销售量的分布等。接下来，我们将采用采样的方式，从上一周的销售数据中获取一部分样本数据。这些样本数据将用于进行非线性规划和 ARIMA 模型的预测与策略制定。通过采样，我们可以减少数据量，提高计算效率，并确保模型能够更好地适应实际情况。然后，我们将使用非线性规划模型制定定价策略。非线性规划模型将综合考虑成本、利润、约束条件等因素，以最大化利润或达到其他目标为目标。我们将根据约束条件，制定合适的约束条件，并使用非线性规划模型求解最佳定价策略。我们还将继续使用 ARIMA 模型来进行销售预测。

（四）问题四的分析

解决问题四，对于考虑更多的因素来更好地制定蔬菜商品的补货和定价决策，我们从节假日期间销售数据分析、促销策略制定、需求预测和库存管理、客流量数据分析、季节性因素考虑、竞争对手价格和策略数据等方面进行分析。

五、模型的建立与求解

（一）数据分析及预处理

数据包含某商超销售的六个蔬菜品类的商品信息和 2020 年 7 月 1 日至 2023 年 6 月 30 日各商品的销售流水明细与批发价格的相关数据以及各商品近期的损耗率数据。

通过整合四个附件的数据到 Excel 数据透视表中，我们将对六个蔬菜品类的日销售量、销售日期、销售单价、销售类型、是否打折以及进价进行统计分析。随后，我们将利用 MATLAB 软件将数据导入，绘制散点图以清洗异常数据。然后，我们将对数据进行 Lilliefors 检验以确定其是否服从正态分布。接着，我们将使用 Q-Q 图和 KS 检验来观察六组数据之间是否存在相似的趋势。进一步，我们将通过蒙特卡罗算法生成随机样本，并计算样本均值。通过统计分析和可视化结果，我们将得到

模拟直方图。最后，我们将使用皮尔斯曼等级相关系数来进一步分析蒙特卡罗分析的结果，以深入了解数据集中变量之间的线性相关程度（见图 1）。

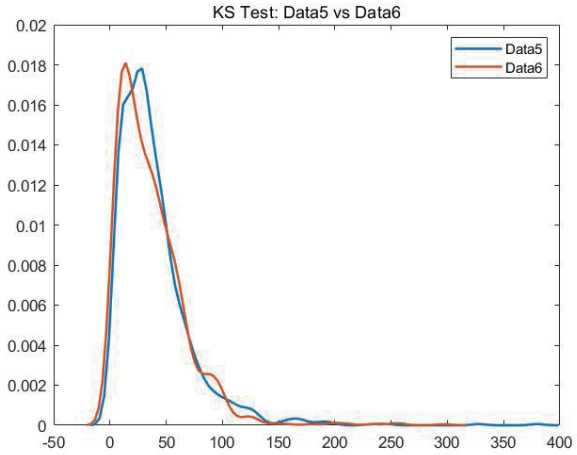

图1 六个蔬菜品类的KS检验

（二）问题一模型的建立与求解

1.Apriori 关联规则算法

Apriori 算法最早由 R.Agrawal 等人提出，它通过逐层搜索的方式迭代寻找数据库中所有项集的关系，以形成规则。该算法被广泛应用于消费市场的价格分析中，通过数据挖掘，商家可以分析商品间的关联关系以预测消费者的购买习惯。该算法首先找出频繁 1- 项集的集合（1L），然后利用 1L 生成候选 2- 项集的集合（2C），并通过支持度筛选获得频繁 2- 项集的集合（2L）。接下来，2L 用于生成候选 3- 项集的集合（3C），再通过支持度筛选生成频繁 3- 项集的集合（3L），如此循环直至不再生成频繁项集为止 [2]。

算法中涉及以下几个重要参数：

（1）支持度

支持度表示某一个项集在全集中出现的可能性

$$support(X \to Y) = P(X,Y) \quad (1)$$

（2）置信度

先决条件 X 发生的前提下，Y 发生的概率

$$Confidence(X \to Y) = P(X|Y) = \frac{P(X,Y)}{P(X)} \quad (2)$$

（3）提升度

前项置信度与后项支持度的比值

$$Lift(X \to Y) = P(X|Y) / P(Y) = \frac{Confidence(X \to Y)}{P(Y)} \quad (3)$$

（4）最小支持度 $(\min\_sup)$

所得到得项目集得支持程度要大于预设得这一值

（5）最小置信度 $(\min\_conf)$

规则在满足最低支持度的情况下，还必须同时满足大于预设好的最小置信度的值

算法步骤如下：

（1）连接：将所有数据进行扫描，得到的全集为第一个集合。而后寻找频繁 $(k+1)$_ 项集 $L_{(k+1)}$，将 k_ 项集合连接回自身产生 $(k+1)$_ 候选集合，记为 C_k。按顺序排列。其中 I_1 和 I_2 是 I_k 的项集，$I_i[j]$ 则表示 I_i 中的第 j 项。进行频繁项集自连接 $L_k \infty L_k$，若 L_k 中前 $(k-1)$ 个项相同：

$$(I_1[1] = I_2[1]) \wedge (I_1[2] = I_2[2]) \wedge \ldots \wedge (I_1[k-1]) = (I_2[k-1]) \wedge (I_1[k]) < I_2[k] \quad (4)$$

则 I_1 和 I_2 可连接。结果为 $I_1[1]I_1[2]...I_1[k-2]I_1[k-1]$

（2）剪枝：L_k 是 C_k 的子集，故 C_k 中元素是否频繁无法确定，但 L_k 的子集全在 C_k 中。重新扫描全集数据，对于 C_k 统计每个元素出现的频次和比重，从而得到支持度 $\text{support}(X \to Y) = P(X,Y)$，剔除支持度过低的项集，以此获得 L_k。重复此循环，当不出现新的频繁集时，结束算法（见图 2）。

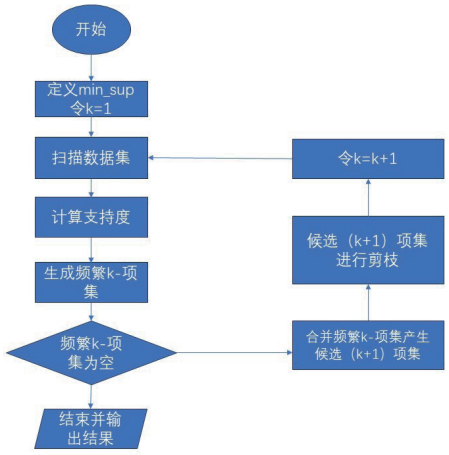

图2 Apriori算法流程图

2. 基于 Apriori 关联规则算法的模型建立

我们将同一天内同一时刻的销售信息归为同一订单，将整个流水信息定义为数据集 D。

$$D = \{t_1, t_2, \ldots\ldots, t_k, \ldots\ldots, t_n\}, t_k = \{i_1, i_2, \ldots\ldots, i_m, \ldots\ldots, i_p\} \quad (5)$$

其中，$t_k(k = 1,2,\ldots\ldots,n)$ 为订单称为事务，$i_m(m = 1,2,\ldots\ldots,n)$ 为订单中购买的数量称为项目。扫描 D，对每个候选进行计数，得到项集 C_1，比较候选支持度计数与最小支持度计数得到项集 L_1，将 $L_1 \infty L_1$ 得到候选项集 C_2，再次扫描 D，对每个候选集进行计数，比较候选支

持度计数与最小支持度计数得到项集 L_2，将 $L_2 \infty L_2$ 产生候选集 C_3……以此重复直到出现 $C_n = \varnothing$，算法结束，找出了所有的频繁项集。再计算所有频繁集中非空子集的最小置信度阈值大于设定值即可输出（见图 3）。

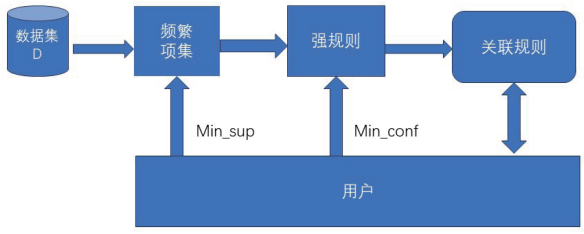

图3 关联规则生成流程图

3. 各品类及单品销售量分布规律和相互关系分析

我们需要导入数据，并将其存储在程序中。通过读取数据文件，我们可以获取单品编码与分类名称和单品名称的对应关系。这样，我们就可以将流水数据与商品信息进行关联。

使用 Apriori 算法，我们可以发现频繁项集和关联规则。通过设置适当的支持度和置信度阈值，我们可以确定哪些商品在消费者的购买行为中经常一起出现，以及它们之间的依赖关系。

首先我们利用统计学算法，对数据进行简单分析与运算，我们可以得到蔬菜各品类销售量情况（见图 4）。

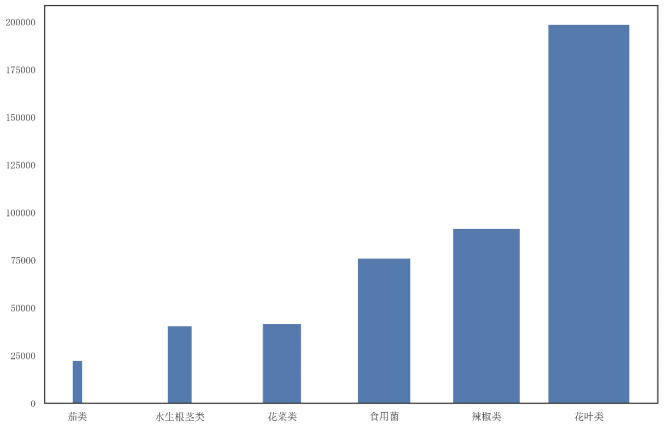

图4 蔬菜各品类销售量情况

花叶类菜品有最大销量，其后依次为食用菌、辣椒类、水生根茎类、茄类和花菜类，我们推断在接下来求解相关关系时，包含花叶类的项集出现频次最高。

我们进一步以月份为划分单位，分析各个蔬菜品类在每个月的销量趋势，得到结果如图 5。

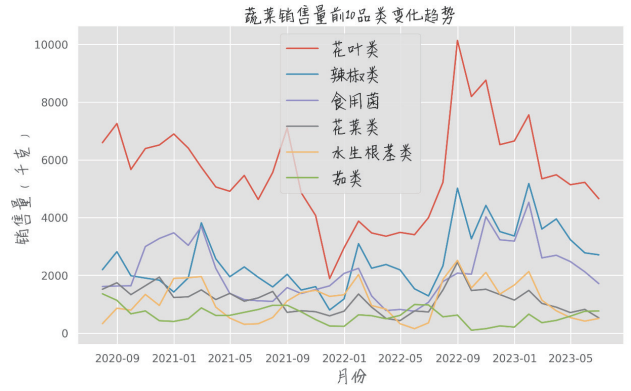

图5 蔬菜各品类销售量变化趋势

从图中可以直观地看到，蔬菜品类销量受时间影响较大，例如花叶类蔬菜在每年的 9 月份都能达到本年度的销售量的最大值；而茄类在下半年往往销量不佳。我们推断在制定补货与定价策略时，时间因素也会是一个重要的考量部分。

而对各个蔬菜单品进行计算与分析后，所得结果空见图 6。

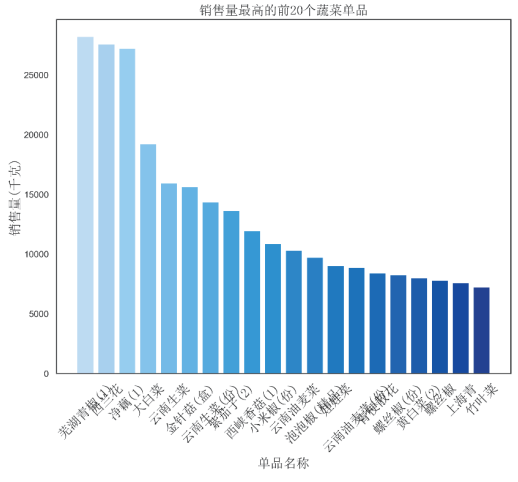

图6 蔬菜单品销售量前十

图中表示了销售最高的前十个蔬菜单品，其中芜湖青椒销量最高，为 28199.151 千克。销量最高的前三个蔬菜品类为芜湖青椒、西兰花、净藕。说明市场对这三种菜品需求量较高，应在补货策略中优先考虑。

同样的，以月份为周期，我们统计计算了蔬菜各频率的平均销售单价及销售总价的变化趋势，计算公式如下：

$$sale\_avg_j = \frac{\sum_i \sum_j sale_{i,j} * sale\_vol_{i,j}}{\sum_i \sum_j sale\_vol_{i,j}} \quad (6)$$

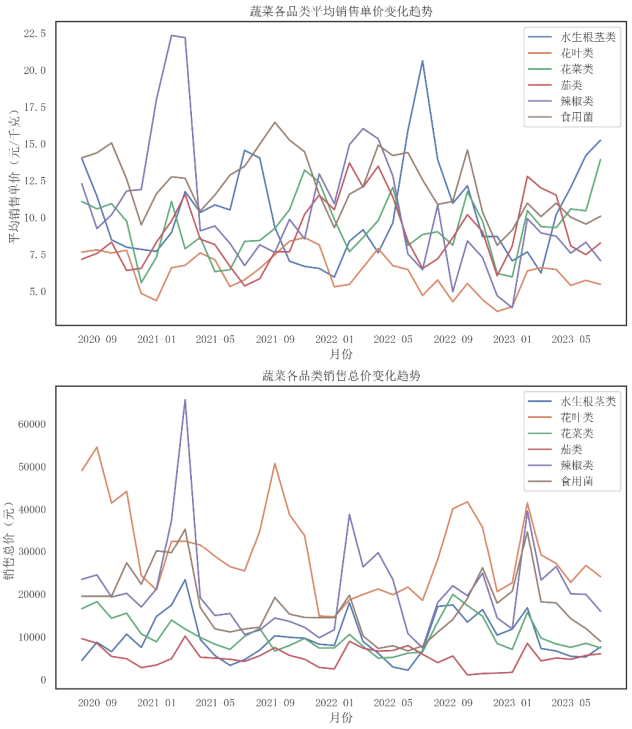

图7 蔬菜各品类单价与总价变化趋势

依据上图可以获取花叶类蔬菜销售单价在每个月基本都是最低的，但其销售总结却非常高，综合销售量，说明单价较低的菜品具有较高的销量，也就是我们常说的“薄利多销”，因而销售量与销售定价间也存在着一定的关系。

综上所述，时间因素对蔬菜销售的影响时不可忽略的，销售量与定价间也存在关系且市场对每种蔬菜的需求量也不同，这些因素都将影响补货与定价策略的制定。

而后，我们将销售流水以时刻为单位进行数据划分，将每时刻作为一个订单项集 I，如此可以避免一天内或多天内的销售流水对相互关系分析造成的干扰。设 $I=\{I_1,I_2,...,I_k\}$，其中 I_k 表示这一订单某一品类或某一单品的销售量。经过一次处理后包含项目的集合用 T 表示，$T\subseteq I$。以其中某一个关联规则举例，该规则为 $X\Rightarrow Y$ 的一个蕴含式，其中 $X\subset I,Y\subset I$，且 $X\cap Y=\varnothing$，其中 X 为先决条件，Y 为关联结果。

关联规则的生成主要依据项集的支持度和置信度，如下：

（1）产生频繁项集（2）产生强关联规则

经多次循环后，所得结果（见图 8、图 9）。

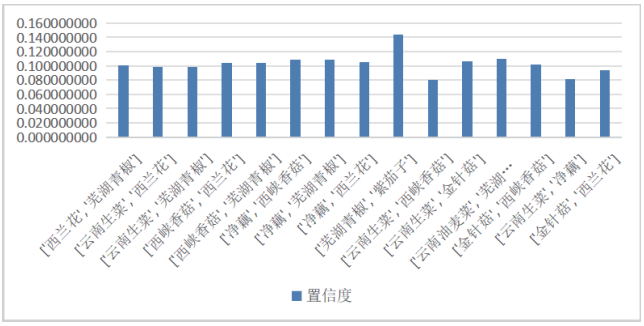

图8 单品频繁项集置信度图

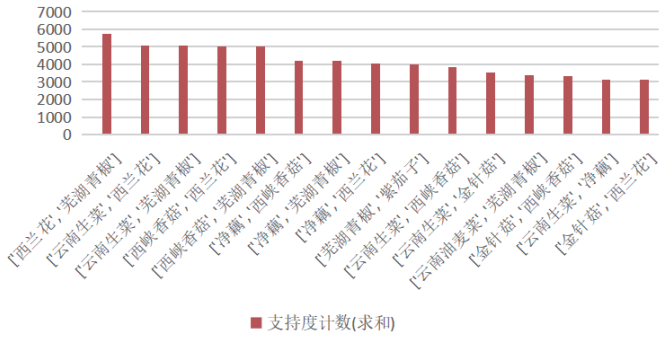

图9 单品频繁项集支持度图

上面的柱形图直观地展示了蔬菜各单品销量之间的置信度和支持度。在频繁项集中，置信度高说明模型对这个组合的输出结果更加肯定；支持度高说明这个组合出现的频次大。综合分析可以得到各单品组合间的相关关系。例如：芜湖青椒与西兰花的相关性较强，紫茄子与芜湖青椒的关联性较强；即当芜湖青椒销量上升时西兰花的销量也会提高。

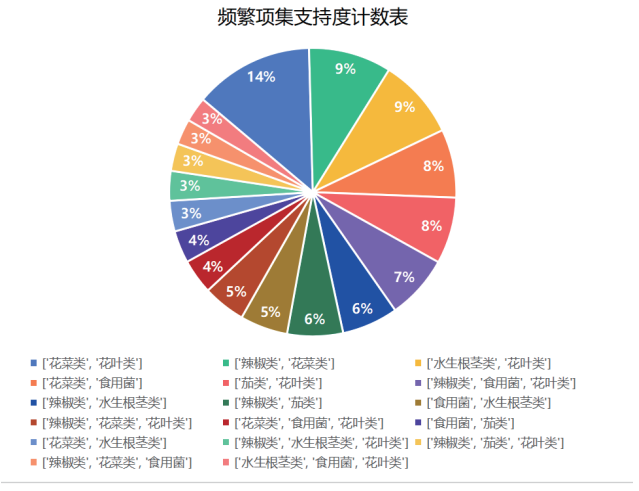

图10 频繁项集支持度图

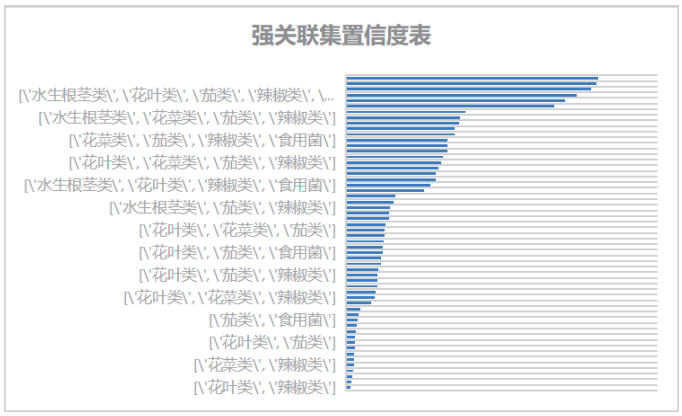

图11 频繁项集置信度图

由上图可以直观看出：

（1）花菜类与花叶类支持度最高为，相关性较强；

（2）辣椒类和花菜类的支持度较高，相关性较强；

（3）当其中一品类销量上升后，另一品类的销量也会随之提升。

4. 皮尔曼等级相关系数模型的建立与预测

同时，我们也应用了皮尔曼等级相关系数模型，对相关性进行分析。皮尔曼等级相关系数是一种用来衡量两个波形之间相似程度的统计指标。当两个波形完全相同时，皮尔曼等级相关系数的值为 1；当两个波形完全相反时，其值为 -1；具体做法如下：

首先，把波形数列 $x=\{x_1,x_2,.....,x_n\}$ 按升序或降序排列得到排序数列 $a=\{a_1,a_2,.....,a_n\}$，将数列 x 内每个元素 x_i 在数列 a 中的位置记为 r_i，称其为元素 x_i 的秩次，从而可以得到数列 x 对应的秩次数列 r。将另一波形数列 $y=\{y_1,y_2,...,y_n\}$ 按同样方式排列得到排序数列 $b=\{b_1,b_2,...,b_n\}$，继而可以得到数列 y 对应的秩次数列 s。将数列 r 和数列 s 内每个元 素对应相减得到秩次差数列 $d=\{d_1,d_2,...,d_n\}$，再将其代入斯皮尔曼等级相关系数公式[3]：

$$\rho=1-\frac{6\sum_{i=1}^{n}d_i^2}{n(n^2-1)} \tag{7}$$

式中：n 为数列点数，对应于一个窗长的采样点数；ρ 为斯皮尔曼等级相关系数。

得到模拟直方图结果见图 12。

5. 两种模型的对比比较

经过我们与模型精度标准的对比，我们发现，皮尔曼等级相关系数模型中得出的结果的精度没有 Apriori 关联规则算法模型得出的结果的精确度高。皮尔曼等级

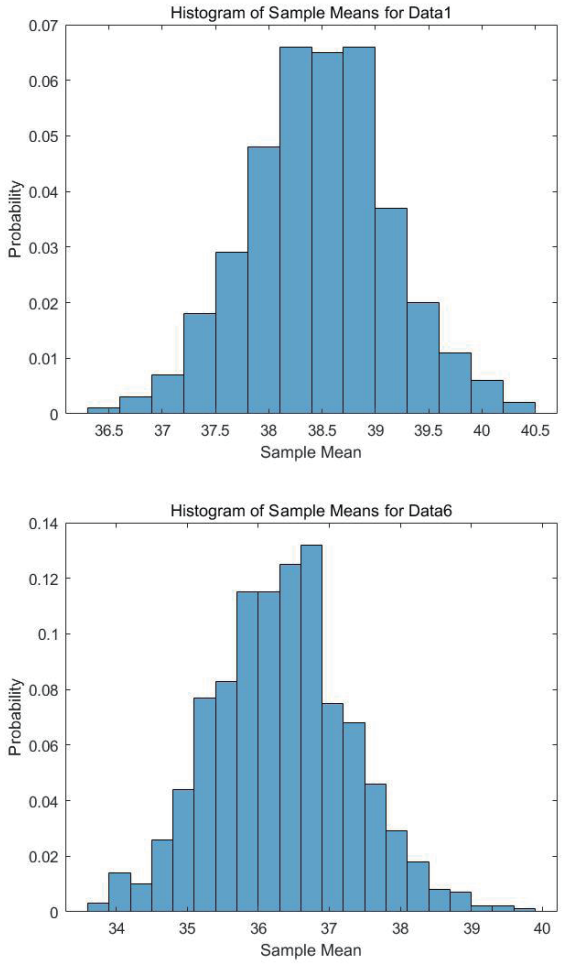

图12 模拟直方图

相关系数忽略了具体数值信息，只基于数据的等级信息进行计算，忽略了具体数值的大小。这意味着在计算过程中，原始数据的数值大小被转化成了等级，部分数据的信息可能会丢失。而 Apriori 算法能够找到数据集中所有频繁项集，不会漏掉任何一个频繁项集。这对挖掘数据集中的潜在关联规非常有用。因此我们决定使用 Apriori 关联规则算法模型。

（二）问题二模型的建立与求解

针对本题，为了制订未来一周的补货与定价策略，使盈利最大化，我们充分结合了第一问的分析结果，考虑了时间、市场需求以及品类与单品间的相关关系对蔬菜销量的影响。而后我们对各蔬菜品类的销售总量和成本加成定价进行分析研讨，求得关系。

1. 基于成本加成定价法和非线性优化 ARIMA 模型的未来一周的补货和定价策略

成本加成定价法，根据是不是新产品 $X=C(1+\omega)$。X 表示价格，C 表示平均成本，ω 表示成本加成率。[4] 该

算法有定价简单，价格稳定的优点，对商超是极其有利的。

在本题中，该定价计算公式如下：

$$P=\frac{\sum sale_{i,j}*sale\_vol_{i,j}}{\sum sale\_vol_{i,j}} \tag{8}$$

为计算出商超所得利润，我们下一步将分析预测蔬菜类的平均批发价格，下为我们对平均批发价的求解公式：

$$whole\_avg\_price_{i,j}=\frac{\sum(1+wastage_j)*sale\_vol_{i,j}*whole\_price_{i,j}}{\sum(1+wastage_j)*sale\_vol_{i,j}} \tag{9}$$

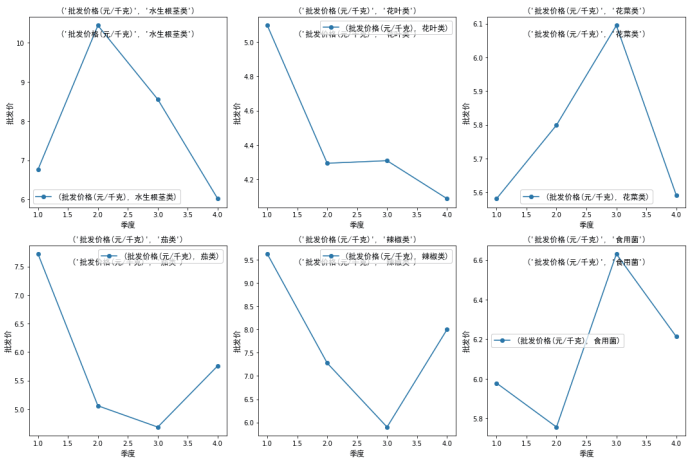

图13 六种品类商品平均批发价格变化

我们以季度为周期进行了数据的分析，分析所得结果如上图所示。

对于未来七天的平均批发价格预测，我们使用了自回归差分移动平均模型（ARIMA），该模型是在时间序列分析中常用的模型，用于描述和预测时间序列数据变化的趋势[5]。$ARIMA(p,q)$ 模型主要结构如下：

$$\begin{cases}\Phi(B)\nabla^d x_t=\Theta(B)\varepsilon_t\\ E(\varepsilon_t)=0, Var(\varepsilon_t)=\sigma_\varepsilon{}^2,\\ E(\varepsilon_t\varepsilon_s)=0, s\neq t; E(x_s\varepsilon_t)=0, \forall s<t\end{cases} \tag{10}$$

其中，$\nabla^d x_t=(1-B)^d x_t=\sum_{i=0}^{d}(-1)^i C_d^i x_{t-i}, \Phi(B)=1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p$ 表示平稳可逆 $ARIMA(p,q)$ 模型的自回归系数多项式，$\Theta(B)=1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q$ 表示平稳可逆 $ARIMA(p,q)$ 模型的移动平滑系数多项式[6]。由此算法可以预测得到未来一周的销售情况，但无法通过此算法直接获得补货和定价策略，所以我们又使用了非线性优化，增加了约束条件。我们使用了不同的函数对蔬菜的平均销售定价与销售总量进行了拟合，首先，根据数据的不同分类，在数据框 avg_wholesale_price_df 中筛选出相应分类的数据，并按照平均销售价格进行排序。使用 linear_func 作为线性函数的拟合函数模型。拟合后结果如表 2 所示。

通过对比分析，确定使用对数函数和线性函数进行拟合。通过分析与计算，最终确定补货销售方案如表 3。

表3 六个品类补货销售策略表

2023/7/1-2323/7/7 补货定价策略													
日期 菜类	花菜类		花叶类		辣椒类		茄类		食用菌		水生根茎类		总进货量
	进货量	定价	进货量	定价	进货量	定价	进货量	定价	进货量	定价	进货量	定价	
2023/7/1	22.098394 93	15.745738 13	158.87939 62	9.5903445 75	76.802550 45	11.084817 69	17.081853 57	13.994031 17	54.471440 81	13.139654 65	10.973772 06	19.617219 59	340.30740 8
2023/7/2	22.090112 89	15.748158 03	158.63878 17	9.6326169 21	76.779079 82	11.094774 83	17.072028 22	14.007935 96	56.415411 65	12.604421 08	10.828608 93	19.699508 9	341.82402 33
2023/7/3	22.088389 91	15.748661 46	158.60799 9	9.6380427 8	76.785376	11.092102 54	17.073471 51	14.005892 46	56.921738 33	12.465014 13	10.813164 9	19.708287 04	342.29013 96
2023/7/4	22.088033 55	15.748765 57	158.60404 95	9.6387392 15	76.783686 02	11.092819 73	17.073259 38	14.006192 78	57.053616 19	12.428704 2	10.811521 09	19.709221 58	342.41416 58
2023/7/5	22.087957 9	15.748787 68	158.60354 26	9.6388285 99	76.784139 56	11.092627 25	17.073290 56	14.006148 64	57.087965 1	12.419246 91	10.811345 09	19.709321 64	342.44824 08
2023/7/6	22.087944 01	15.748791 74	158.60347 75	9.6388400 8	76.784017 84	11.092678 91	17.073285 98	14.006155 13	57.096911 62	12.416783 67	10.811320 78	19.709335 5	342.45695 78
2023/7/7	22.087938 41	15.748793 38	158.60346 92	9.6388415 52	76.784050 51	11.092665 04	17.073286 65	14.006154 17	57.096241 83	12.416809 09	10.811315 86	19.709338 3	342.45930 25

表2 拟合模型参数表

分类名称	模型名称	模型参数
水生根茎类	对数函数	[-26.11036 2.51727 83.60512]
花叶类	对数函数	[-36.00776 2.35435 209.75696]
花菜类	线性函数	[-2.89165 64.20212]
茄类	对数函数	[-8.16177 1.62367 36.47012]
辣椒类	对数函数	[-17.0756 3.10817 105.1638]
食用菌	线性函数	[-3.28878 92.53733]

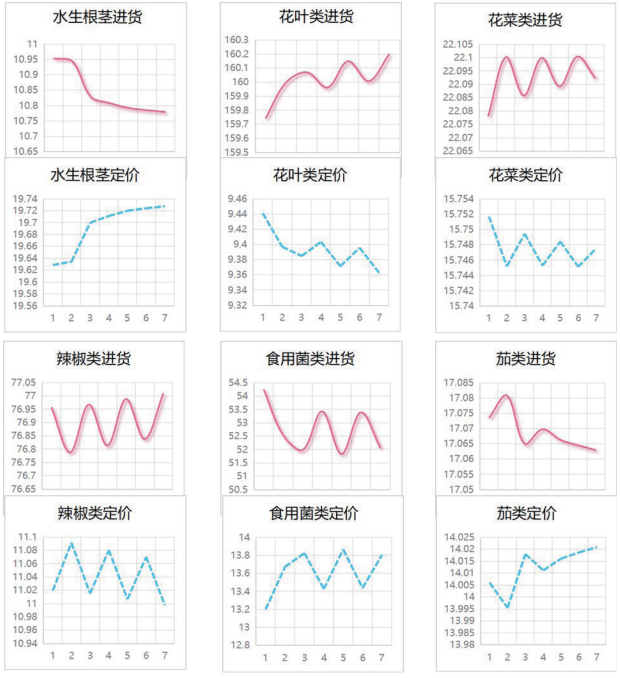

图14 六个品类补货销售策略图

2. 梯度提升决策树算法的未来一周的补货和定价策略

梯度提升是一种常用的训练方式。其中，梯度提升决策树是利用梯度提升训练得到的决策树模型，每轮训练得到一个弱学习器，通过拟合残差得到一组弱学习器，即强分类器[7]。

在该模型中，计算总收益的算法如下：

$$P\_ = \sum (sale_{i,j} * sale\_vol_{i,j} - \cos t_j) \tag{11}$$

使用的损失函数如下：

$$L(y_i, f(x_i)) = \frac{1}{2}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{12}$$

其中，$\hat{y}_i$表示预测值，y_i表示实际值，$\hat{y}_i - y_i$则为预测值与实际值之间的差值，即梯度。在训练中，对每个树节点的分裂采用贪心算法，即在每个分裂都选择增益最大的样本特征值，具体计算方式如下：

$$Gain = \frac{1}{2}[\frac{G_L^{\ 2}}{H_L^{\ 2}+\lambda} + \frac{G_R^{\ 2}}{H_R^{\ 2}+\lambda} - \frac{(G_L+G_R)^2}{(H_L+H_R)^2+\lambda}] - \gamma \tag{13}$$

其中，G_L是左边所有样本特征值排序后的一阶梯度和，G_R是右边所有样本特征值排序后的一阶梯度和，H_L是左边所有样本特征值排序后的二阶梯度和，H_R是右边所有样本特征值排序后的二阶梯度和，γ为控制常量。

3. 两种模型的对比

在实际求解问题过程中，发现梯度提升决策树算法存在以下几个问题：容易产生过拟合，对异常值与噪声过度敏感，推广性不强；涉及参数过多，难以调优。故放弃此算法改用非线性优化 ARIMA 模型。

（四）问题三模型的建立与求解

1. 基于遗传算法的未来一天的补货和定价策略

在第二问的基础上，控制可售单品数以及保持单品订购量满足最小陈列量要求已然成为制订补货策略时要考虑的必要因素。为做到最大满足市场需求的情况下，同时保证商超收益最大，我们假设本年度 7 月 1 日需要补货的单品与前一周存在相同趋势，因此在下图我们展示出上一周的蔬菜单品销售情况，合计 49 种可售菜品。

在此问中我们为了使结果更加接近实际，采用了遗传算法，遗传算法是一种通过模拟自然进化过程来 搜索最优解的优化算法。其基本原理是将问题 的解表示为一个染色体，然后通过选择、交叉和 变异等操作来优化染色体[8]。

我们首先确定了种群中个体数量与编码方式，将每一种策略表示为一个染色体，染色体由若干基因构成，每个基因代表一个单品的补货量与销售价的策略。初始种群可以通过随机生成若干种策略得到。通过选择、交叉、变异、繁殖等多种操作，以可售单品总数控制在 27-33 个，且各单品订购量满足最小陈列量 2.5 千克为约束条件，进行迭代计算。

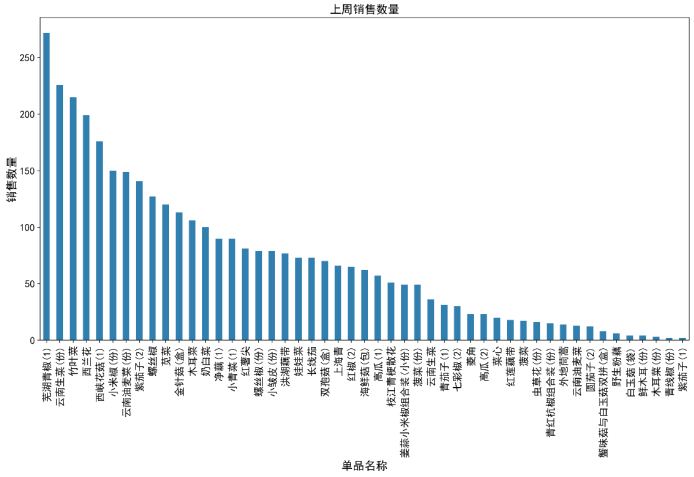

图15 7月1日上一周销售数量图

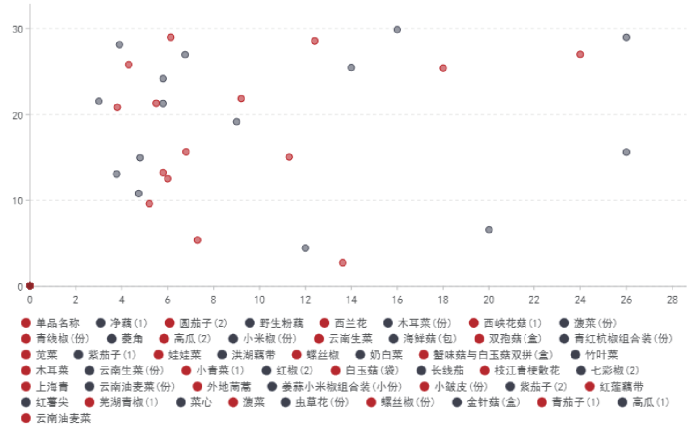

图16 各单品补货销售策略图

所得该商超2023年7月1日的单品补货与定价策略如表4所示。

（五）问题四模型的建立与求解

问题四提出，为更好制定补货与定价策略，还应增加哪些因素的考量。

商超在节假日期间的销售数据分析对于制定补货和定价策略至关重要。研究表明，节假日是消费者冲动性购买和集中购买的高峰时期[3]。因此，商超应当积极采集节假日期间菜品的销售数据，并进行深入分析。首先，通过收集节假日的销售数据，商超可以了解特定节假日的蔬菜销售趋势和消费者购买行为。这些数据可以揭示一些有价值的信息，例如，消费者在特定节假日期间对于不同种类蔬菜的需求变化、消费者偏好的蔬菜品种以及购买决策的主要因素等。商超可以根据这些数据，调整补货计划，增加库存量以足节假日期间的需求。其次，商超可以利用节假日销售数据来制定节假日促销策略，以吸引顾客。例如，在销售数据分析的基础上，商超可

表4 各单品补货销售策略表

单品名称	进货量	销售价格
净藕 (1)	29.82	16.00
圆茄子 (2)	28.94	6.13
野生粉藕	28.93	26.00
西兰花	28.52	12.41
木耳菜 (份)	28.08	3.90
西峡花菇 (1)	26.95	24.00
菠菜 (份)	26.91	6.76
青线椒 (份)	25.75	4.30
菱角	25.40	14.00
高瓜 (2)	25.35	18.00
小米椒 (份)	24.13	5.80
云南生菜	21.80	9.20
海鲜菇 (包)	21.50	3.00
双孢菇 (盒)	21.25	5.50
青红杭椒组合装 (份)	21.23	5.80
苋菜	20.80	3.81
紫茄子 (1)	19.11	9.00
娃娃菜	15.61	6.80
洪湖藕带	15.57	26.00
螺丝椒	15.01	11.29
奶白菜	14.92	4.79
蟹味菇与白玉菇双拼 (盒)	13.19	5.80
竹叶菜	13.02	3.77
木耳菜	12.49	6.00
云南生菜 (份)	10.75	4.74
小青菜 (1)	9.57	5.20
红椒 (2)	6.53	20.00
白玉菇 (袋)	5.33	7.30
长线茄	4.40	12.00
枝江青梗散花	2.69	13.63
七彩椒 (2)	0.00	0.00
上海青	0.00	0.00
云南油麦菜 (份)	0.00	0.00
外地茼蒿	0.00	0.00
姜蒜小米椒组合装 (小份)	0.00	0.00
小皱皮 (份)	0.00	0.00
紫茄子 (2)	0.00	0.00
红莲藕带	0.00	0.00
红薯尖	0.00	0.00
芜湖青椒 (1)	0.00	0.00
菜心	0.00	0.00
菠菜	0.00	0.00
虫草花 (份)	0.00	0.00
螺丝椒 (份)	0.00	0.00
金针菇 (盒)	0.00	0.00
青茄子 (1)	0.00	0.00
高瓜 (1)	0.00	0.00
云南油麦菜	0.00	0.00

以针对特定节假日推出折扣动、组织特色品展示或举办烹饪示范活动等。这些促销策略可以吸引更多顾客，增加节假日期间的销售额。此外，商超还可以利用节假日销售数据来进行需求预测和库存管理。通过分析历史的节假日销售数据，商超可以预测未来节假日期间的蔬菜需求量，从而合理安排补货计划，避免库存过剩或供应不足的问题。这样做不仅可以提高销售效益，还能节约成本，确保蔬菜供应的稳定性。

季节性因素也是商超应该重视的因素。各类蔬菜在不同的季节受欢迎程度可能存在差异，而对于这些季节性的变化，商超需要及时调整补货策略，以满足市场需求。商超应当兼顾考虑不同季节的蔬菜需求量、销售量和价格数据变化，力求精准把握供给需求，调整采购计划，确保库存充足。

竞争对手的价格和策略数据也是商超需要关注的重要因素。假设该商超存在多个竞争者，当竞争者的价格印象低于（或高于）该商超时，应变动一定比例的价格高于（或低于）竞争者价格，该模型可以使用如下函数进行表示：

$$\omega_{h,d}=\sum_{r=1}^{k}\tau_{h,r}\omega_{h,r}(1-\delta_{d,r}) \tag{13}$$

$$\delta_{d,r}=\sum_{h=1}^{g}(\omega_{h,r}/\omega_{h,d})(q_{h,r}/\sum_{h=1}^{g}q_{h,d})-1 \tag{14}$$

利用以上公式可以对竞争者价格印象$\delta_{d,r}$进行计算，并划分结果来判断竞争者的价格印象水平。我们定义其划分标准如表 5 所示。

表5 经营竞争者划分标准

价格印象δ值	竞争者
$\delta_{d,r}<-0.04$	低价格印象
$-0.04\le\delta_{d,r}\le0.04$	中价格印象
$\delta_{d,r}>0.04$	高价格印象

结合分析，我们得出了以下结论：

1. 对于消费者需求量大的品类，其对价格也较敏感，故应慎重定价且充分重视低价格印象竞争者。

2. 对于消费者需求和价格敏感度都不高的品类，定价时应考虑中等印象竞争者。

3. 对于需求量最小的品类，消费者大都因为方便而选择就近采购，可以参考高价格印象竞争者。

除了这些关键数据，商超还应该关注库存状况、天气和节假日数据以及供应链和物流数据。了解当前各种蔬菜的生产及库存情况可以帮助商超提前判断哪些蔬菜可能会因为库存冗余导致新鲜度下降而需要打折销售，或者哪些蔬菜可能因为短期购买量增多导致供不应求而需要调高价格。天气和节假日数据对于预测消费者的购物行为至关重要。例如，在天气寒冷的冬季，消费者更倾向于减少购买频率而增加单次购买量，而在节假日期间，消费者通常也购买更多的食材来准备餐饮和聚会。最后，供应链和物流数据可以帮助商超预测未来的库存状况，及时调整补货策略调整定价方案，以确保可以及时、可靠地对商品供应进行补充。

综上所述，收集客流量数据、季节性因素数据、竞争对手数据、消费者反馈数据、库存状况数据、天气和节假日数据以及供应链和物流数据对商超的补货和定价决策至关重要。这些数据可以提供关键的市场情报和运营指导，帮助商超优化经营效益、提高客户满意度，并确保商超在竞争激烈的市场中保持领先地位。

六、评价与改进

（一）模型的评价

1. 模型的优点

（1）Apriori 算法模型

①简单易懂：Apriori 算法的基本原理直观且易于理解。它基于频繁项集的生成和剪枝过程，通过扫描数据集来识别频繁项集和关联规则。

②可扩展性强：Apriori 算法能够处理大规模数据集，适用于大部分实际应用场景。它通过利用候选项集的自连接和剪枝操作减少计算和存储开销，有效提高算法的效率。

③能够找到所有频繁项集：Apriori 算法能够找到数据集中的所有频繁项集，不会漏掉任何一个频繁项集。这对于挖掘数据集中的潜在关联规则非常有用。

④灵活性：Apriori 算法具有良好的灵活性，可以通过调整最小支持度和最小置信度等参数控制频繁项集和关联规则的挖掘程度。

⑤可解释性强：Apriori 算法生成的关联规则易于解释和理解。它可以提供直观的关联规则，帮助人们发现数据集中的有用关联性，并进行决策和推荐。

（2）ARIMA 模型

模型简单，非常灵活，可以建立不同时序数据的预测模型，不需要借助外部变量，可以有多种选择。

（3）非线性优化

能将复杂的非线性问题进行处理，非线性优化可以

应用于各种实际问题，包括经济学、工程学、物理学等领域中的非线性模型。它能够处理目标函数和约束条件非线性的情况，提供了更广泛的建模和求解能力；可以找到全局或近似最优解；可以添加约束条件，非线性优化算法能够考虑约束条件，如等式约束和不等式约束，以确保求解结果满足问题的限制条件。

2. 模型的缺点

（1）Apriori 算法模型处理大规模数据集效率低下，对数据稀疏性敏感。

（2)ARIMA 模型要求时序数据为稳定的，只能处理线性关系，不能处理非线性关系，无法对周期性、季节性的数据进行预测。

（3）非线性优化过度依赖于初始选择的算法和初始点。非线性优化算法的性能和结果很大程度上依赖于选择合适的算法和初始点。不同的算法可能对不同类型的问题表现出不同的效果，而初始点的选择可能会影响算法的收敛速度和结果质量；容易陷入局部最优解；计算复杂度过高，非线性优化问题的求解通常需要进行多次迭代计算，计算复杂度较高。尤其是对于大规模问题，计算时间和内存消耗会成为问题。

（二）模型的推广

该模型提供了基于数据分析的补货和定价策略，旨在最大化商超的利润并减小损失。然而，该模型的有效性和可行性受到数据的质量和数量等因素的重大影响。在使用之前，需要对数据进行大量的预处理工作，以确保数据的准确性和足够的覆盖范围。此外，该模型还受到题中所设定的假设条件的限制，如市场稳定性和需求与价格成正比等。因此，在实际应用中，需要进一步优化和调整该模型，以适应实际情况和需求。这可能涉及考虑更多的因素，如竞争情况、市场趋势和牌策略等，并根据实际效果进行反和改进。总之，该模型提供了一种有用的基于数据分析的方法，但需要经过充分的数据预处理和实情况的优化与调整，以确保其在实际应用中的有效性和可行性。

参考文献：

[1] 艾媒咨询 . 深度解读 2023-2024 年中国生鲜电商运行大数据及发展前景研究报告 [EB/OL].(2023-06-09).
[2] 郭艳萍 , 高云 , 景雯 . 基于 Apriori 算法的大气污染物关联性分析研究 [J]. 软件工程 ,2023,26(9):8-11.
[3] 张梦琦 . 基于 Apriori 算法的关联规则分析 [D]. 大连 : 大连理工大学 ,2022.
[4] 赵韩 . 基于 RNN 的时序数据多步预测方法的研究与应用 [D]. 北京 : 北京工业大学 ,2020.
[5] 贾科 , 杨哲 , 魏超等 . 基于斯皮尔曼等级相关系数的新能源送出线路纵联保护 [J]. 电力系统自动化 ,2020,44(15):103-111.
[6] 安琪 . 基于成本加成定价法的科技查新服务定价研究 [J]. 图书馆研究与工作 ,2021(10):25-31.
[7] 吴会会 , 王嘉鹏 , 吴文静等 . 基于 ARIMA 模型的全球气表温度预测分析 [J]. 现代信息科技 ,2023,7(16):147-150.
[8] 周跃 . 基于遗传算法的机器人路径规划优化研究 [J]. 电子元器件与信息技术 ,2023,7(6):65-68.
[9] 黄梅 . 基于可信执行环境的公平隐私联邦梯度提升决策树系统研究 [D]. 桂林 : 广西师范大学 ,2023.
[10] 张清华 . 零售商场节假日促销方法探讨 [J]. 商场现代化 ,2016(8):30-31.
[11] 方莉 . 高校后勤蔬菜原料价格制定方法研究 [J]. 高校后勤研究 ,2021(10):22-24.