

# 基于混合预测模型的供应链商家商品需求研究

王雨佳 薛梓玥 闫诗梦  
北京工商大学数学与统计学院  
DOI:10.12238/acair.v2i2.7379

**[摘要]** 本文基于现有销售数据,对供应链商品的售卖规律、未来各时期销售量进行建模分析。首先针对商品数据进行相关描述性统计分析;分别使用Croston间歇性时间序列预测和基于粒子群算法的ARIMA-SVR混合预测模型对各商品需求进行预测及比较分析,得到较为准确的预测结果,数据准确率较高。选择K-means方法将数据分成10类。再从数据中分别挑选与每个序列相同的时间构成等长序列;对数据中的等长序列使用余弦相似度进行分析,计算序列间的相似度,找出最相似的序列;然后分别使用之前构建的序列的ARIMA-SVR混合预测模型对与其相似的序列进行预测分析,再对结果进行的指标评价,发现数据准确率较高。最后,本文对所使用的预测与分类模型进行分析,发现模型精度较高。

**[关键词]** Croston间歇时间序列预测;基于PSO的ARIMA-SVR混合预测模型;K-means聚类;余弦相似度

中图分类号: O642.5+45 文献标识码: A

## Research on commodity demand of supply chain merchants based on hybrid forecasting model

Yujia Wang Ziyue Xue Shimeng Yan

School of Mathematics and Statistics, Beijing Technology and Business University

**[Abstract]** Based on the existing sales data, this paper models and analyzes the sales rules of supply chain commodities and the sales volume in the future. Firstly, the relevant descriptive statistical analysis of commodity data is carried out, and the Croston intermittent time series prediction and ARIMA-SVR hybrid prediction model based on particle swarm optimization are used to predict and compare the demand of each commodity, and the more accurate prediction results are obtained, and the data accuracy is high. The K-means method was chosen to divide the data into 10 categories. Then, the same time as each series was selected from the data to form an equal-length series, the cosine similarity of the equal-length series in the data was used to analyze the similarity between the sequences, and the similarity between the sequences was calculated to find the most similar sequences. Finally, this paper analyzes the prediction and classification models used, and finds that the model accuracy is high.

**[Key words]** Croston intermittent time series forecasting; ARIMA-SVR hybrid prediction model based on PSO; K-means clustering; Cosine similarity

### 1 介绍

随着互联网+时代快速发展,电子商务行业快速发展<sup>[1]</sup>。电商平台承载了数以千计的商家,商家将各种商品存放在电商平台的各个仓库中<sup>[2]</sup>。电商平台面临着巨大的供应链管理挑战,要通过科学的管理手段和智能决策来提高供应链的效率。供应链优化不仅可以显著降低库存成本,还能够保证商品按时履约,提供卓越的客户服<sup>[3]</sup>。

### 2 模型建立及求解

本文基于电商零售商家的相关历史数据开展研究,先进行数据预处理,生成一个包括所选择时间范围内所有日期的日期

序列。对于没有出现的日期,将需求设置为0。最后将各个商家在各个仓库的各种商品的需求序列分离出来,以便后续预测分类。

指数平滑法是以平滑系数对时间序列的历史数据按时间先后顺序给予不同的权重,时间越近,权重越大,时间越小,权重越小,且权重的减小是以等比级数递减的。指数平滑法具有保留全部历史数据、算法简单、结果稳定的特点,能对间断性物资进行有效的预测。

$$y_{t+1} = \alpha y_t + (1 - \alpha)y_t$$

Croston预测法和指数平滑法都是预测间断性物资的方

法, Croston将需求间隔  $k_t$  与需求量  $d_t$  分开, 采用指数平滑法分别计算需求间隔和需求量: 若发生需求, 则更新需求间隔和需求量的估计, 然后将需求量除以需求间隔得到平均需求, 用平均需求来预测; 若不发生需求则保持原来的预测, 只更新从上次发生需求到现在的需求间隔。计算公式如下式(2)(3)所示, 其中,  $0 \leq \alpha, \beta \leq 1$ 。

$$\hat{x}_{t+1} = \begin{cases} \hat{x}_t & \text{if } d_t = 0 \\ (1-\alpha)\hat{x}_t + \alpha d_t & \text{if } d_t \leq 0 \end{cases} \quad (2)$$

$$\hat{k}_{t+1} = \begin{cases} \hat{k}_t & \text{if } d_t = 0 \\ (1-\beta)\hat{k}_t + \beta k_t & \text{if } d_t \leq 0 \end{cases} \quad (3)$$

ADI和CV是目前常用的两种表示间歇性需求的特征指标, 其计算公式分别如(4)-(6)所示。ADI表示某条序列中的平均需求间隔, 其中,  $n_i$  表示第  $i$  条序列有  $n_i$  个周期,  $z_i$  表示第  $i$  条序列中非零周期的数量。CV表示单条序列的变异系数。  $s_i$  表示非零需求周期的标准差, 其中,  $l$  表示非零需求周期的个数,  $x_z^{(i)}$  表示第  $i$  个非零需求周期,  $\bar{x}_z$  表示非零需求周期的平均值。

$$ADI = n_i / z_i \quad (4)$$

$$CV = S_i / \bar{x}_z \quad (5)$$

$$S_i = \sqrt{\frac{1}{l} \sum_{i=1}^l (x_z^{(i)} - \bar{x}_z)^2} \quad (6)$$

此外, ADI和CV能很好地描述间歇性需求数据的特点, 在实际业务中常用来对序列分类, 需求模式随机出现, 大量时间段没有需求且各个时期的需求之间差异较大, 伴随有大量零需求阶段。考虑利用ADI和CV对这些商家、仓库、商品形成的序列进行分类, 使得同一类别在需求上的特征最为相似。预测准确率  $wmape$  是平均绝对百分比误差的加权, 用于对不同数据点的重要性进行加权考虑。  $wmape$  的值越低, 预测的准确性越高。它用于衡量关注不同数据点的重要性时的预测准确性, 定义如下:

$$1 - wmape = 1 - \frac{\sum |y_i - \hat{y}_i|}{\sum y_i} \quad (7)$$

其中,  $y_i$  为第  $i$  个序列(商家在各仓库中存放的各种商品每天的数量)的真实需求量,  $\hat{y}_i$  为第  $i$  个序列的预测需求量。  $1-wmape$  的值越接近1说明预测效果越好。

采用间歇性需求时间序列预测Croston方法对各商品需求量进行预测, 部分预测结果如下表1所示, 预测准确率指标的值与1的偏离度很大, 预测效果不好。数据基本符合正态分布, 模型的预测性能还好, 但存在一定的偏差。再采用ARIMA模型对各商

家在各仓库的商品的需求量进行预测, 使用PSO算法来调整和优化ARIMA和SVR模型的参数。PSO算法用于搜索最佳参数组合, 以最大程度地减小预测误差。使用历史数据对模型进行训练, 并结合PSO算法来不断调整模型的参数。使用ARIMA、SVR、优化预测模型对各商家在各仓库的商品的需求量进行预测, 结果的精度对比, 如表2所示。

表1 Croston部分预测结果表

product_no	seller_no	warehouse	ADI	CV	S	1-wmape
product_27	seller_5	wh_2	3	0.57	5.19	-0.056
product_27	seller_5	wh_6	2.19	0.58	4.49	0.097
...	...	...	...	...	...	...
product_30	seller_5	wh_7	1.31	0.70	6.15	0
product_31	seller_5	wh_1	2	0.88	83	0.452

表2 预测模型评估精度表

指标名称	RMSE	MAPE	R <sup>2</sup>
ARIMA	4.909	3.798	0.761
SVR	12.2287	9.1901	0.8732
优化预测模型	4.412	3.7742	0.9292

由表2可知, 优化预测模型的R<sup>2</sup>为0.92, 其余指标较小, 误差较小, 模型的拟合效果较好。由此, 这种结合ARIMA、SVR和PSO的方法可以帮助提高商品需求预测的准确性, 特别是需要考虑多因素的复杂需求预测问题中。

K-means是把样本集作为输入, 对样本实行聚类过程, 将拥有相似特征的数据样本聚成一类, 类别内的样本点尽量紧密集中在一起, 而类别间的距离差异尽可能的明显, 聚类结果则越理想。定义一个D维的欧几里得空间中的数据集  $X = \{x_1, x_2, \dots, x_n\}$ , 集群中心  $Z = \{z_1, z_2, \dots, z_k\}$ ,  $y = [y_{ic}]_{n \times k}$  用来指示数据点  $x_i$  是否属于第  $c$  集群, 其中,  $y_{ic} \in \{0, 1\}$ ,  $c \in [1, k]$

采用轮廓系数、DBI、CH等聚类评估指标用来评估聚类算法的性能和质量。轮廓系数是所有样本轮廓系数的均值, 取值范围是[-1, 1], 同类别样本距离越相近不同类别样本距离越远, 分数越高, 聚类效果越好。DBI用来衡量任意两个簇的簇内距离与簇间距离之比。该指标越小表示聚类效果越好。CH通过计算类内各点与类中心的距离平方和来度量类内的紧密度, 通过计算类间中心点与数据集中心点距离平方和来度量数据集的分离度, CH指标由分离度与紧密度的比值得到, CH越大表示聚类效果越好。

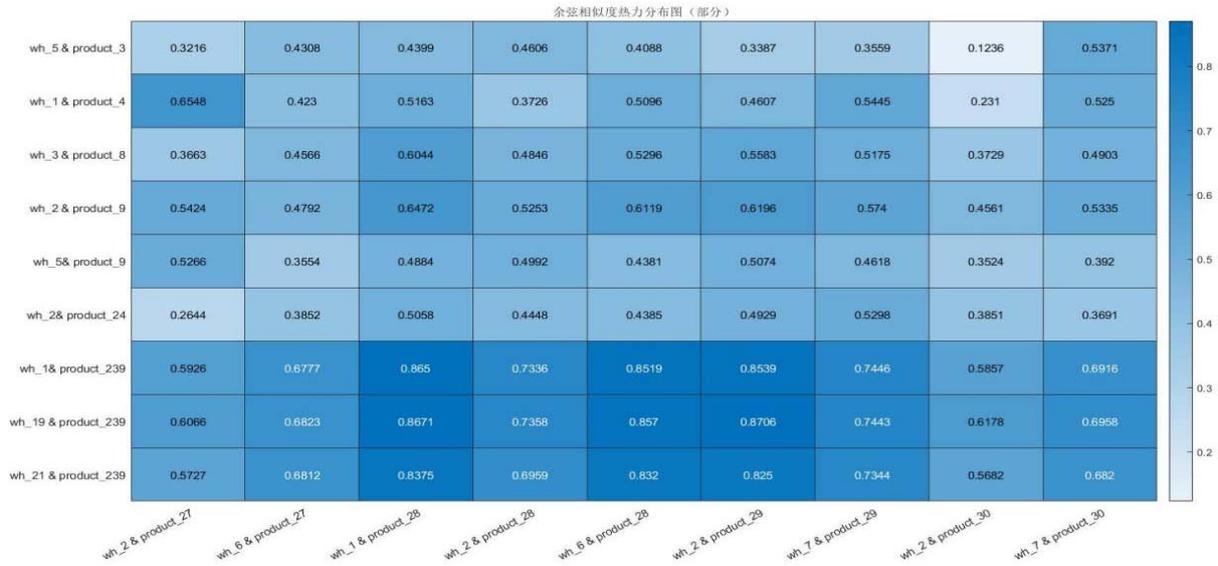


图1 余弦相似度热力分布图

K-means聚类将商家、仓库和商品划分为K个簇,每个簇代表相似的销售需求模式。将各个时间序列划分为10类,代表10种相似的销售需求模式。聚类评估指标结果如下表3所示。可知聚类效果较好。

表3 聚类评估指标结果表

Contour Coefficient	DBI	CH
0.804	0.126	105.641

针对可能由于商品新上市而出现的新数据进行分类。在上文对历史数据利用k-means算法进行了分类。此处考虑计算新出现的序列与历史序列的相关性,进而找到新出现的序列的相似序列。再根据各个序列特点代入对应训练好的预测模型中进行预测。

余弦相似度通过计算两个夹角的余弦值来评估它们之间的相似度。将向量根据坐标值绘制到向量空间中,计算两向量的夹角  $\theta$ ,通过  $\theta$  的大小去判断两向量的相似程度。夹角越小说明越相似,反之不相似。假设存在两个n维向量  $A$  和  $A = [A_1, A_2, \dots, A_n]$ ,  $B = [B_1, B_2, \dots, B_n]$ ,余弦值总是在  $[-1, 1]$  之间,余弦值越趋近于1,代表两向量方向越接近;越趋近于-1,代表两向量方向越相反;接近于0,代表两个向量近乎正交。一般情况下,需要将相似度归一化到  $[0, 1]$  区间内,如式(8)所示。

$$COSSIM = \cos\theta * 0.5 + 0.5 \quad (8)$$

先对序列进行数据预处理。由于新出现的序列的时间非同一个时间段,序列并不等长,要根据新出现的序列找历史序列对应的天数再进行余弦相似度的计算,再将相似度较高的序列归为同一类序列。由于数据量过大,选取部分余弦相似度计算结果做可视化呈现,部分余弦相似度计算结果如图1所示。

由上图可知,哪些序列之间相似度最高。由此,对这些新序列代入训练好的特定的优化预测模型进行预测,输出新数据未

来的需求趋势。上文已经训练好了历史序列对应的需求量预测模型,根据计算出的余弦相似度对其进行分类,将相似序列输入上文建好的特定预测模型中预测,部分结果如下表4所示。

表4 部分新序列预测结果表

seller_no	product_no	warehouse_no	date	forecast_qty
seller_19	product_2215	wh_21	2023/5/16	6
seller_19	product_2215	wh_21	2023/5/17	6
...	...	...	...	...
seller_19	product_407	wh_23	2023/5/19	14
seller_19	product_407	wh_23	2023/5/20	14

### 3 总结

本文建立时间序列预测模型,用现有数据分析供应链商品的销售模式并预测未来的销售量。此外,对未来的时间序列数据进行可靠性评估,有助于评估预测结果的可信度。但由于长时间序列数据存在较多的不确定性,导致该模型的预测效果易受到外部因素的影响,在长期预测方面往往效果不佳。在本研究基础上,还可扩展以提高预测供应链商品销售模式的准确性和实用性。综上,未来的供应链商品销售预测将更加智能、准确和适应性强,以更好地满足供应链商家的需求。

#### [参考文献]

- [1]郎祎平.面向工业大数据的时间序列迁移预测方法[D].河南:河南师范大学,2022.
- [2]智能供应链:预测算法理论与实践[M].北京:电子工业出版社,2023.
- [3]郎祎平,毛文涛,罗铁军,等.间歇性时间序列的可预测性评估及联合预测方法[J].计算机应用,2022,42(9):2722-2731.