Identifying Popular Products at An Early Stage of Sales Season for Apparel Industry

Jiayun Wang

School of Management
Zhejiang University

Sep 2023

Table of Contents

Introduction

- Traditional rule-based method is both time-consuming and error-prone.
- Research on classifying the popularity of apparel products is limited.

Sell-through Rate	Days on Market					
	<15 days	>= 15 days	>= 30days	>= 45 days	>= 60 days	>= 90 days
< 10%	——	Slow	Slow	Slow	Slow	Slow
>= 10%	——	Slow	Slow	Slow	Slow	Slow
>= 15%	——	Slow	Slow	Slow	Slow	Slow
>= 20%	Average	Slow	Slow	Slow	Slow	Slow
>= 25%	Average	Average	Slow	Slow	Slow	Slow
>= 35%	Fast	Average	Average	Slow	Slow	Slow
>= 45%	Fast	Fast	Average	Average	Slow	Slow
>= 55%	Fast	Fast	Fast	Average	Slow	Slow
>= 65%	Fast	Fast	Fast	Fast	Average	Slow
>= 75%	Fast	Fast	Fast	Fast	Average	Average
>= 80%	Fast	Fast	Fast	Fast	Fast	Fast

Figure: Rule-based Method

Framework

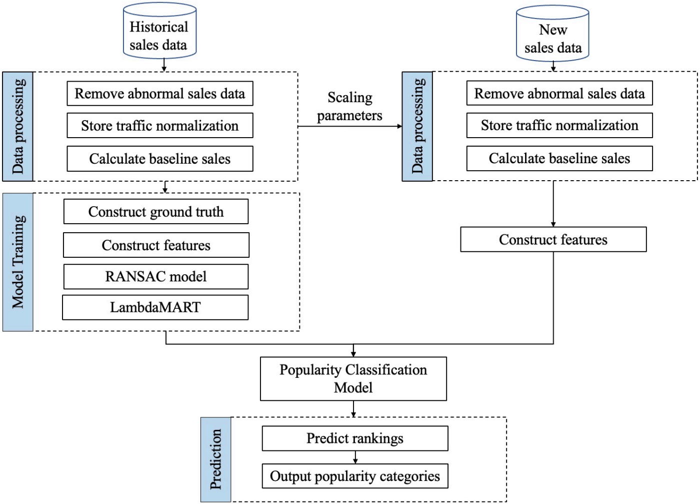

Figure: Framework of the Popularity Classification Model

Data Processing

Data cleaning

Remove sales that are not from individual buyers, including group sales data and returned sales data.

Normalize store traffic

Adjust sales from different stores by normalizing the store traffic in each region to a standardized level.

Baseline sales

Calculate the baseline sales volume using linear regression, which estimates the sales volume that would have occurred if no discounts were applied to the products.

Dependent Variable: AW Sales

Definition of AW Sales

AW Sales is calculated as the store's average weekly sales during its main sales period within a given region, using adjusted sales data.

Main sales period identification

- Calculate the percentage of sales for each week.
- Sort the weekly sales percentages in descending order.
- Compute the cumulative percentage of weekly sales in descending order.
- Identify the weeks in which the cumulative percentage exceeds the threshold as the main selling period.
- If the number of weeks in the main selling period is less than the preset minimum number of weeks, additional weeks are included until the minimum is reached.

Independent Variable: Feature Description.

Table 5 Mathematical Notations

Notation	Definition
k	The index of product.
K	Number of products to be ranked.
N_k	Number of stores with an initial stock of product k.
M_k	Number of weeks since the launch of product k.
s_{mnk}	The adjusted sales volume of product k in store n during week m.
x_{mnk}	A binary variable: 1 if store n sells any quantity of product k in week m, and 0 otherwise.
$\mathcal{D}_{nk}$	$\mathcal{D}_{nk}$ is a sequence of days that represent the ordering times of orders of product k in store n. As an illustration, for product k in store n, suppose that we record sales on day 1 (1 order), day 5 (3 orders), and day 7 (2 orders) after its initial launching. The vector $\mathcal{D}_{nk}$ would be denoted as (1, 5, 5, 5, 7, 7). Therefore, $\mathcal{D}_{nk}[0]$ represents the number of days before the first sale of product k in store n since its launching. Subsequently, the difference between $\mathcal{D}_{nk}[d+1]$ and $\mathcal{D}_{nk}[d]$ represents the number of days that elapsed between two consecutive orders of product k in store n.
d_{nk}	Length of the vector $\mathcal{D}_{nk}$.
$\mathcal{L}_k$	$\mathcal{L}_k$ refers to the longest increasing subsequence derived from the sequence of "weekly adjusted sales volume" of product k across all stores within a typical region. To illustrate, consider product k with a weekly adjusted sales volume sequence of (50, 162, 364, 285, 398, 488) in a region. In this case, $\mathcal{L}_k$ would be (50, 162, 364, 398, 488).
l_k	Length of $\mathcal{L}_k$.

Note. Here, $\mathcal{D}_{nk}[d]$ and $\mathcal{L}_k[d]$ denote the $(d+1)$th elements of $\mathcal{D}_{nk}$ and $\mathcal{L}_k$, respectively.

Figure: Mathematical Notations

Table 6 Definitions of Variables Used in the Popularity Classification Model

Variable	Definition
Mean sales	$\sum_{n=1}^{N_k} \frac{\sum_{m=1}^{M_k} s_{mnk}}{M_k} \cdot \frac{1}{N_k}$
Mean maximal sales	$\sum_{n=1}^{N_k} \max_m (s_{mnk}) \cdot \frac{1}{N_k}$
Active store ratio	$\sum_{n=1}^{N_k} \frac{\sum_{m=1}^{M_k} x_{mnk}}{M_k} \cdot \frac{1}{N_k}$
First sale waiting days	$\sum_{n=1}^{N_k} \mathcal{D}_{nk}[0] \cdot \frac{1}{N_k}$
Mean waiting days	$\sum_{n=1}^{N_k} \sum_{d=0}^{d_{nk}-1} \frac{\mathcal{D}_{nk}[d+1] - \mathcal{D}_{nk}[d]}{d_{nk}-1} \cdot \frac{1}{N_k}$
Increasing subsequence length (Length)	l_k
Increasing subsequence range (Range)	$\mathcal{L}_k[l_k - 1] - \mathcal{L}_k[0]$
Increasing subsequence growth rate (Growth Rate)	$\frac{\mathcal{L}_k[l_k - 1] - \mathcal{L}_k[0]}{l_k}$

Figure: Definitions of Variables

Prediction Model: Ranking-based Algorithm.

Evaluation metric

$$DCG = \sum_{i=1}^{K} \frac{2^{aws_i} - 1}{log(1 + i)} \quad (1)$$

$$NDCG = \frac{DCG}{IDCG} \quad (2)$$

In our context, K represents the number of products to be ranked, while aws_i denotes the value of *AW Sales* of the product at position i.
The term *IDCG* refers to the Ideal Discounted Cumulative Gain, which represents the DCG value of the ideal order. The ideal order can be obtained from the *AW Sales* calculated based on the entire sales season.

Prediction Algorithm

Table 2 Precision and Recall Rates of 6 Different Models in Identification of Fast-selling Products

	Algorithm	14 Days	21 Days	1 Month	2 Month	3 Month
Precision	LambdaMart	**0.635** ⋆	0.663	**0.733** ⋆	**0.841** †	**0.895** †
	MLR	0.632	0.569	0.604	0.759	**0.770** ⋆
	SVM	**0.720** †	**0.729** †	**0.762** †	0.530	0.573
	XGB	0.594	**0.680** ⋆	0.717	0.747	0.706
	RF	0.557	0.576	0.604	0.632	0.661
	NN	0.611	0.648	0.724	**0.766** ⋆	0.732
Recall	LambdaMart	**0.794** †	**0.794** †	**0.809** †	**0.779** †	**0.784** †
	MLR	0.629	**0.634** ⋆	**0.673** ⋆	**0.733** ⋆	**0.776** ⋆
	SVM	0.267	0.347	0.396	0.614	0.574
	XGB	0.564	0.515	0.351	**0.733** ⋆	0.713
	RF	**0.728** ⋆	0.564	0.619	0.535	0.589
	NN	0.708	0.520	0.559	0.728	0.721

Notes. The red numbers († sign) represent the largest value among seven models and the blue numbers (⋆ sign) represent the second largest value.

Figure: Comparision with other machine learning methods

Model Evaluation

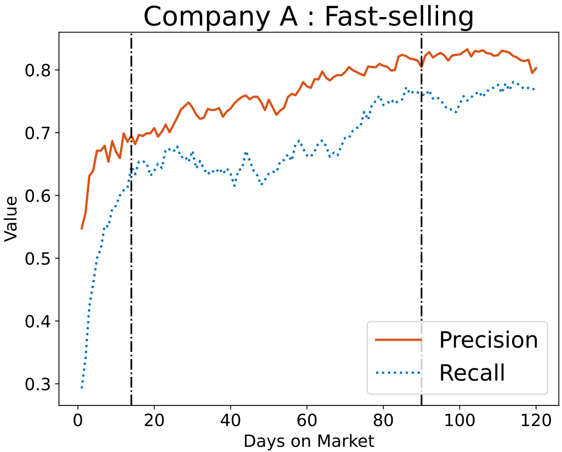

Figure: Precision and Recall Rates at Different Days on Market

Implementation

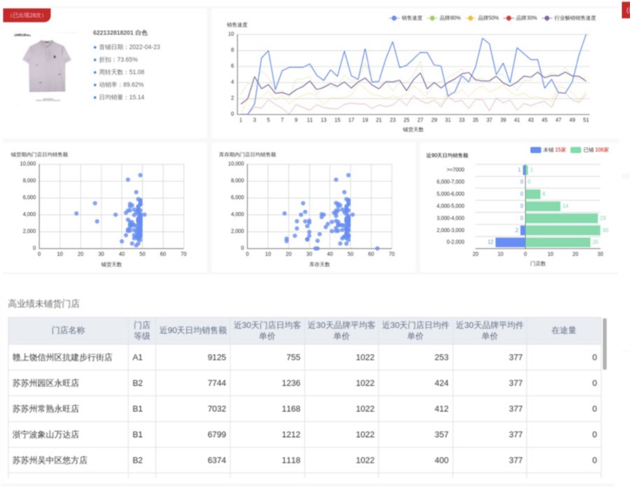

Figure: UI of develped software module

Decrease in the number of orders.

The number of products with more than one order decreases by 66% compared to the same period in 2021. The maximum number of orders placed on one product drops from 7 to 5.

Increase in sell-through rate.

Increase in sell-through rate from 50.88% to 85.53% at the end of the season in 2021 and a 74.4% reduction in disposal costs.

Increase in sales volume.(assumption)

If the sales managers place orders completely according to the model's predictions, the sales volume can be improved by 39,558 units, accounting for 5.9% of actual total sale volume.

Conclusion

Paper Overview

- Introduces a novel approach to identify fast-selling apparel products in the early stages of the selling season using machine learning techniques.
- Proposes the AW Sales metric to mitigate biases inherent in traditional metrics.
- Constructs innovative features, such as the longest increasing subsequence, to capture product sales trends and abnormal sales volumes.

Publication

Informs Journal on Applied Analytics(Interface)

Jiayun Wang, Shanshan Wu, Qingwei Jin, Yijun Wang, and Can Chen. Identifying Popular Products at An Early Stage of Sales Season for Apparel Industry.
https://doi.org/10.1287/inte.2023.0022