Pricing the Long Tail by Explainable Product Aggregation and Monotonic Bandits

Marco Mussi, Gianmarco Genalti, Francesco Trovò, Alessandro Nuara, Nicola Gatti and Marcello Restelli

{marco.mussi, gianmarco.genalti, francesco1.trovo, nicola.gatti, marcello.restelli}@polimi.it, alessandro.nuara@mlcube.com

POLITECNICO MILANO 1863

ML cube

What is the Long Tail?

Long tail [Anderson, 2006] consists in sell:

- a small number of products with high volumes
- a large number of products with low volumes

Motivation

Why This Work?

- An increasing number of e-commerce are joining the long-tail paradigm
- No other work exploits the peculiarities of the long-tail framework to make dynamic pricing

Where We Start?

- We work with an e-commerce to price over 20000 products:
  - ≈ 1000 are best seller
  - ≈ 12000 are long tail with at least a sale
  - ≈ 7000 have never been sold
- The market presents seasonalities and trends

Which are the Challenges?

- Design a learning algorithm that is *robust* and *sample efficient*:
  - Robust: essential when data are scarce and noisy, as in real-world settings
  - Sample efficient: essential in non-stationary settings to limit delay in learning
- Find an effective solution to cluster products:
  - We cannot rely only on transaction data (too scarce)
  - Long-tail products have different market dynamics than best-seller → trivial one-to-one aggregations may fail

Setting and Goal

Setting

- We have a textual description and transaction data for product $j \in \mathcal{J}$ ($\mathcal{J}$ is the set containing all the products)
- At every time t, we aim to set a margin $m_{jt} := \frac{p_{jt} - c_j}{c_j}$ (p_{jt} and c_j are the selling price and the acquisition cost)
- $v_{jt}(m_{jt})$ is the actual number of sales (volumes) for an item j at time t when choosing margin m_{jt}

Goal

Select the margin maximizing the *total profit*:

$$m^*_{jt} = \arg\max_{m_{jt} \in \mathcal{M}_j} f_{jt}(m_{jt})$$

$$f_{jt}(m_{jt}) := m_{jt}\, c_j\, v_{jt}(m_{jt})$$

Algorithm

Distance Estimation

Idea: Compute a *distance matrix* using textual descriptions

- The distance is computed for **all the couples** of products using **Term Frequency-Inverse Document Frequency** (TF-IDF) algorithm
- We obtain a matrix $\mathcal{D} = [d_{jk}]_{j,k \in \mathcal{J}}$, where d_{jk} is the distance between any couple of items $j, k \in \mathcal{J}$

Tree Structure Generation

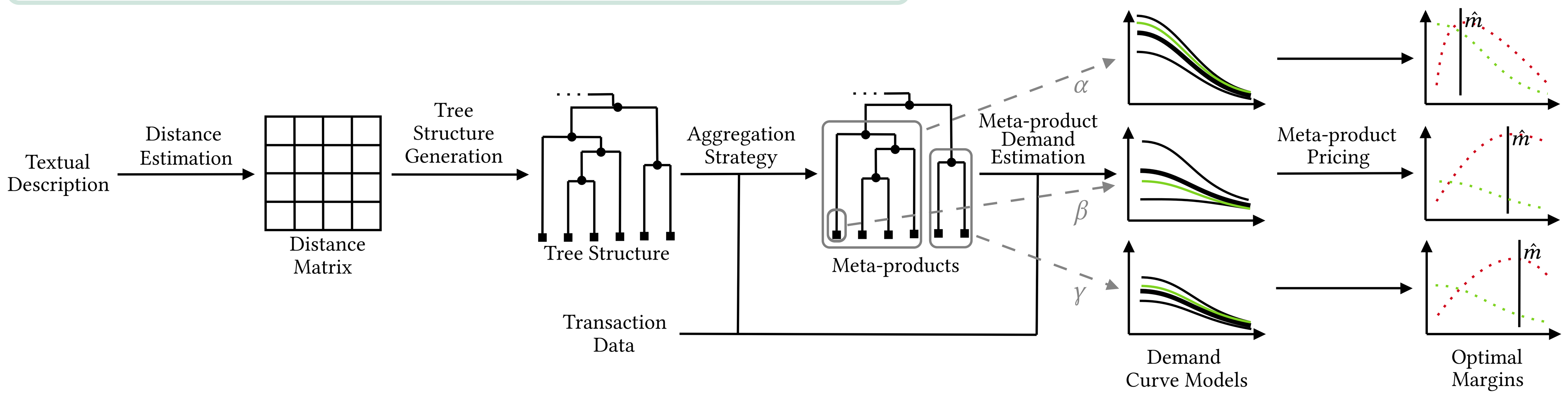
Idea: Generate a *binary tree* from the distance matrix $\mathcal{D}$

- In this tree:
  - Each **leaf** represents a product $j \in \mathcal{J}$
  - Each **non-terminal** node represents a **meta-product** $\mathcal{K}$

Aggregation Strategy

Idea: Map every *product* to a *meta-product*

- Return a set of **minimal meta-products**, each with a sufficient amount of data to get an **accurate estimate of the demand** curve
- Ensuring every selected meta-product is (the **minimal**) provided by at least a given percentage of non-zero (aggregated) volumes samples



Meta-product Pricing

Idea: Price product j using meta-product $\mathcal{K}$

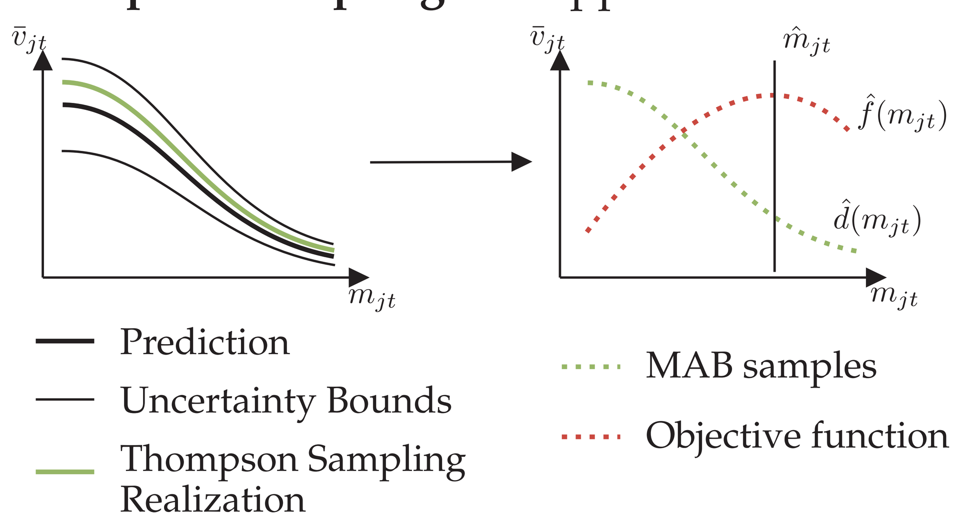
- Meta-product is composed by products $k \in \mathcal{K}$
- **Volumes** of products $k \in \mathcal{K}$ are **aggregated** and **deseasonalized**

$$\bar{v}_{\mathcal{K}\tau} := s_{\mathcal{K}\tau} \sum_{k \in \mathcal{K}} v_{k\tau}, \quad \forall \tau \in \mathcal{T}$$

- **Margins** of products $k \in \mathcal{K}$ are **averaged**, considering **volumes as weights**

$$m_{\mathcal{K}\tau} := \sum_{k \in \mathcal{K}} m_{k\tau} \cdot \frac{v_{k\tau}}{\sum_{h \in \mathcal{K}} v_{h\tau}}, \quad \forall \tau \in \mathcal{T}$$

- Optimal margins are selected according to a **Thompson Sampling**-like approach



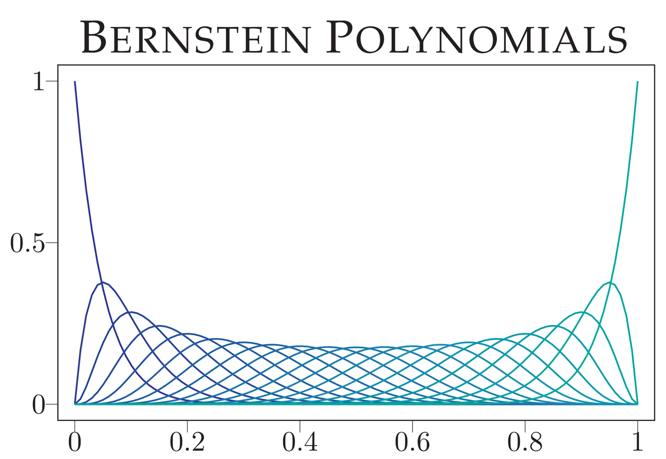
Demand Curve Models

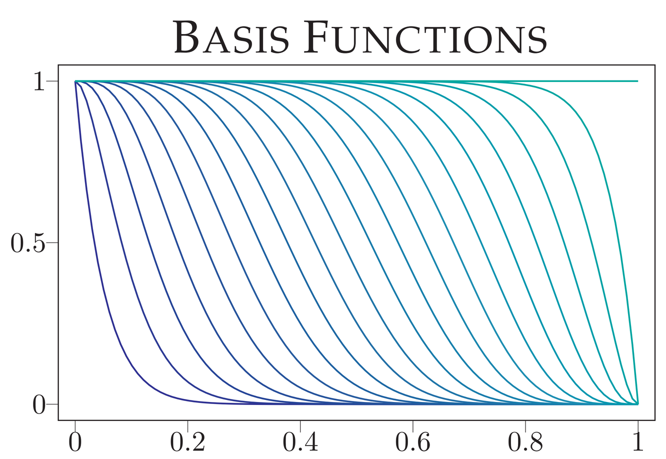
Idea: Model the non-increasing monotonicity *margin* → *volumes*

- The model considers a **subset** of historical data, related to the last N weeks → **trend** effects are **negligible**
- The **seasonality** is stable → **removed** using historical data
- The demand is modeled as $\hat{d}_j(m) = \sum_{h=0}^{Z} \theta_h\, \phi_h(m)$
- The demand curve is forced to be **monotonic non-increasing** with a **Bayesian Regression** Model → *LogNormal* prior θ_h over all the basis functions $\phi_h(m)$
- The basis functions are ($h = \{0, \dots, Z\}$):

$$\phi_h(m) = B_Z(m) \cdot (I_{Z+1} - S_{Z+1})^{-1} \cdot \mathbb{I}_h$$

where $B_Z(m)$ is the vector of **Bernstein Polynomials** (degree Z) and S_{Z+1} is the square matrix with all 1 in the *superdiagonal*

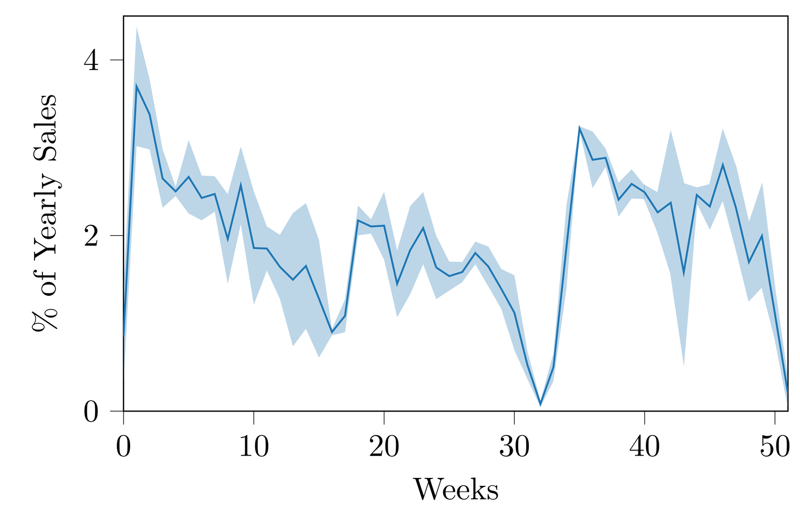




Real-World Application

Real-World Setting

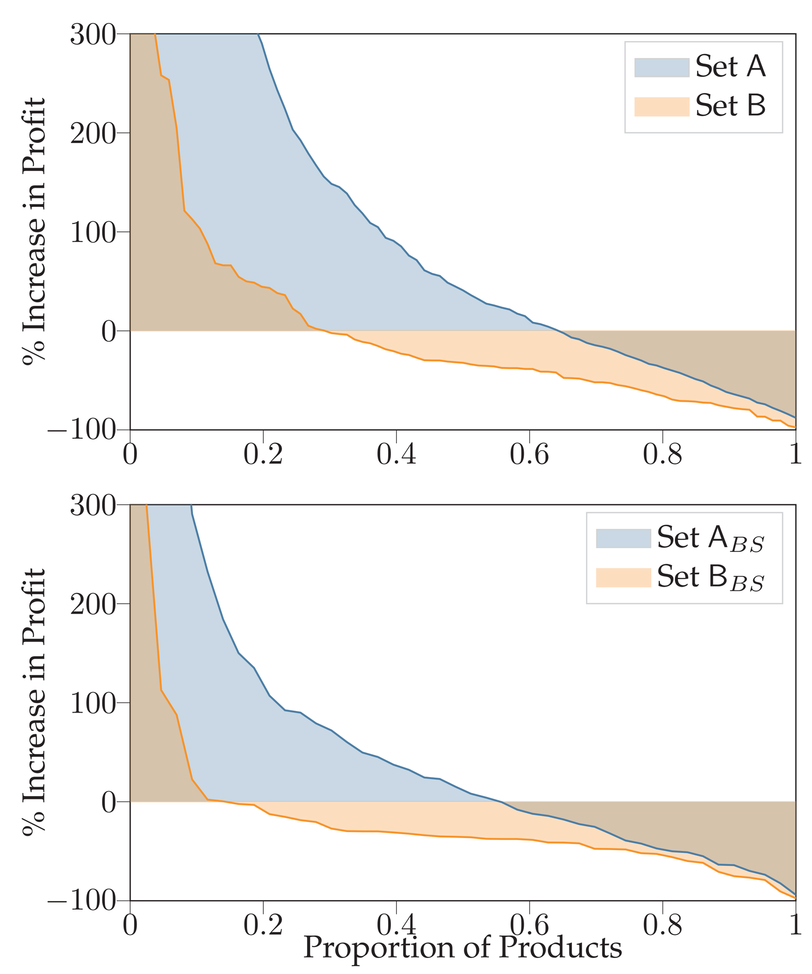
- A/B test involving ≈ 8000 products with
  - ≈ 2.5 MEuros of turnover
  - ≈ 0.5 MEuros of margin
- The test includes both long-tail and best-seller products
- The test is conducted for 8 weeks in fall/winter 2021
- The performances are matched with the ones of set B, considering as control period the same time-span of the previuos year



Results Overview

Best-Seller	Long-Tail	Overall
+18%	+91%	+40%

Variations in Products' Margin



References

Chris Anderson. *The long tail: Why the future of business is selling less of more.* Hachette Books, 2006.

Hamsa Bastani, David Simchi-Levi, and Ruihao Zhu. Meta dynamic pricing: Transfer learning across experiments. *Management Science*, 2022.

Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *FOCS*. IEEE, 2003.