Demand Estimation of Full-Cut Promotion on E-Commerce Company

WONG JING HUI
SUPERVISOR: ASST PROF YAN ZHEN ZHEN

CONTENTS

WHAT IS FULL-CUT PROMOTION?

Pay minimum threshold amount to enjoy fixed discount to entire order.

discount amount

Threshold amount

Figure 1. Extracted from VIPshop.com

Figure 2. Extracted from Taobao.com

Figure 3. Extracted from Tmall.com

OBJECTIVE

Motivations for researching on full-cut promotions in E-commerce.

Growing popularity

Full-cut promotions are popular and widely used in E-commerce such as TaoBao and VIPshop.

No existing literature

No existing literature on the demand estimation and optimisation of full-cut promotions in E-commerce.

Costly promotions

Promotions are very costly for an E-commerce company. We seek to implement revenue management of full-cut promotions.

Demand Estimation + Optimisation

Our objective is to introduce an approach that estimates customers' demand towards full-cut promotion, optimize revenue from the promotion.

OBJECTIVE

Challenges in Demand Estimation of full-cut promotions.

Full-cut promotion sensitivity

In an aggregated transactional dataset, no information on which customers are subscribed to full-cut promotion. How do we identify the types of full-cut promotion? How do we measure their sensitivity to full-cut promotion?

Customer Data Sparsity

Customer segmentation often involve using customer demographics. However, due to privacy issues, the data is not readily available. Also, highly sparse. How can we segment customers well using a product-level transactional dataset?

METHODOLOGY

Outline of methodology.

Exploratory Data Analysis

Identify the types of Full-cut promotions. Understand the distribution of discount and threshold amounts.

Propose model for promotion sensitivity

Define promotion sensitivity, differentiate customers' attraction towards full-cut promotion using on price-to-threshold distance, no. of products purchased.

Customer Segmentation

Segment customers based on promotion-driven demand. Comparative study of choice probabilities with Conditional Gradient approach.

Machine Learning + Optimisation

Apply machine learning to investigate relationship between features and promotion demand. Set up framework to optimize threshold and discount amount for revenue management.

EXPLORATORY DATA ANALYSIS

Identify discount amount.

Figure 4. Histogram plot for full-cut discount amount in each order

Figure 5. Histogram plot for discount amount in each order (¥100-¥200 interval)

Full-Cut Discount Amount	Threshold Amount
¥100	¥199
¥200	¥398
¥300	¥597

Table 1. Proposed Full-cut discount amount and threshold amount

EXPLORATORY DATA ANALYSIS

Identify corresponding threshold amount.

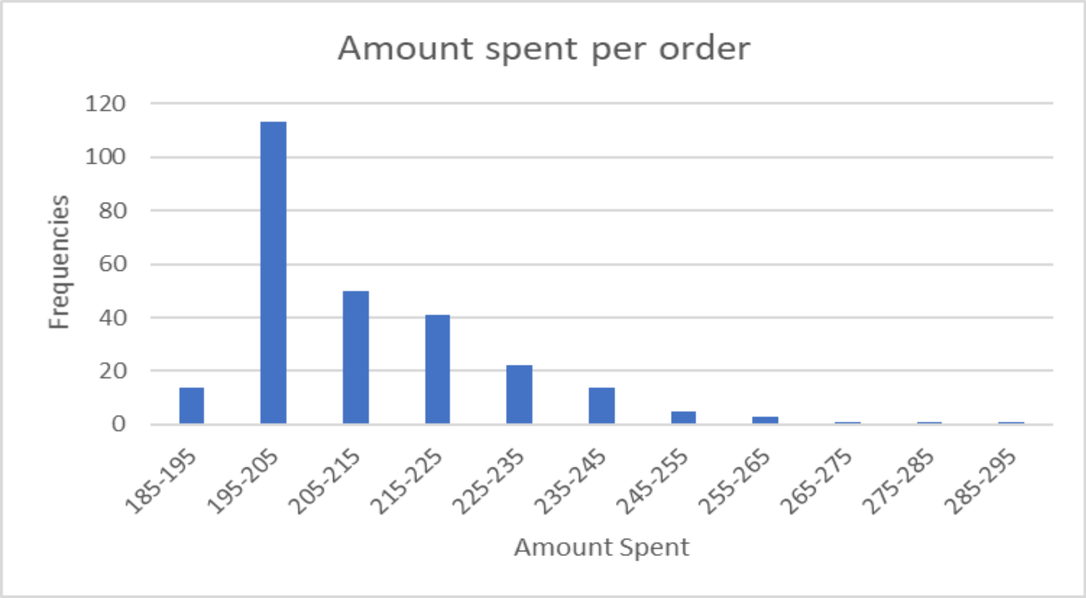


Figure 6 Histogram plot for amount spent per order (for ¥100 discount amount)

Full-Cut Discount Amount	Threshold Amount
¥100	¥199
¥200	¥398
¥300	¥597

Table 1. Proposed Full-cut discount amount and threshold amount

Figure 7. Histogram plot for amount spent per order (for ¥200 discount amount)

Figure 8. Histogram plot for amount spent per order (for ¥300 discount amount)

PROPOSED MODEL FOR PROMOTION SENSITIVITY

Formulation of promotion-driven demand on threshold and order amount.

$$\mathcal{Y}_i = \begin{cases} \max\{0, \log(\frac{Threshold_j/Price_i}{1 - Threshold_j/Price_i})\}, & Price_i \geq Threshold_j, \mathcal{T} \geq 2 \\ 0, & Price_i < Threshold_j \ or \ \mathcal{T} < 2 \end{cases}$$

$$\boldsymbol{Target_i} = \begin{cases} 1, & Price_i = Threshold_j, \mathcal{T} \geq 2 \\ \frac{\mathcal{Y}_i - \min(\mathcal{Y})}{\max(\mathcal{Y}) - \min(\mathcal{Y})}, & Price_i > Threshold_j, \mathcal{T} \geq 2 \\ 0, & Price_i < Threshold_j or \ \mathcal{T} < 2 \end{cases}$$

Logit transformation is represented as $\log\left(\frac{p(X)}{1-p(X)}\right)$.

$Threshold_j/Price_i$ represents $p(X)$, the probability of a customer being attracted to the full-cut promotion for $Price_i \geq Threshold_j$

CUSTOMER SEGMENTATION

Clustering based on promotion-driven demand.

	Optimal number of clusters	Mean Silhouette score
K-means	3	0.6367
K-medoids	3	0.6209
Hierarchical clustering (Ward Linkage)	3	0.5842
Hierarchical clustering (Ward Linkage)	4	0.6201
Hierarchical clustering (Single Linkage)	3	0.6557
Hierarchical clustering (Complete Linkage)	4	0.6011
Hierarchical clustering (Average Linkage)	3	0.6139
Hierarchical clustering (Average Linkage)	4	0.6018

Figure 9. Summary of clustering models used

CUSTOMER SEGMENTATION

K-means selected due to silhouette score and interpretability of clusters.

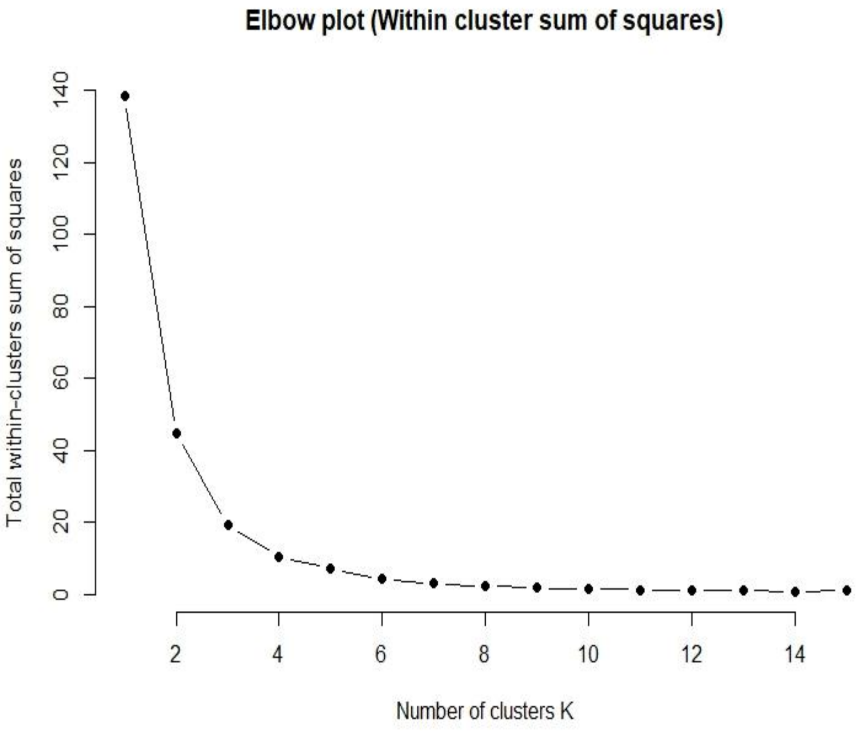


Figure 10. Elbow plot to determine number of clusters (K-means)

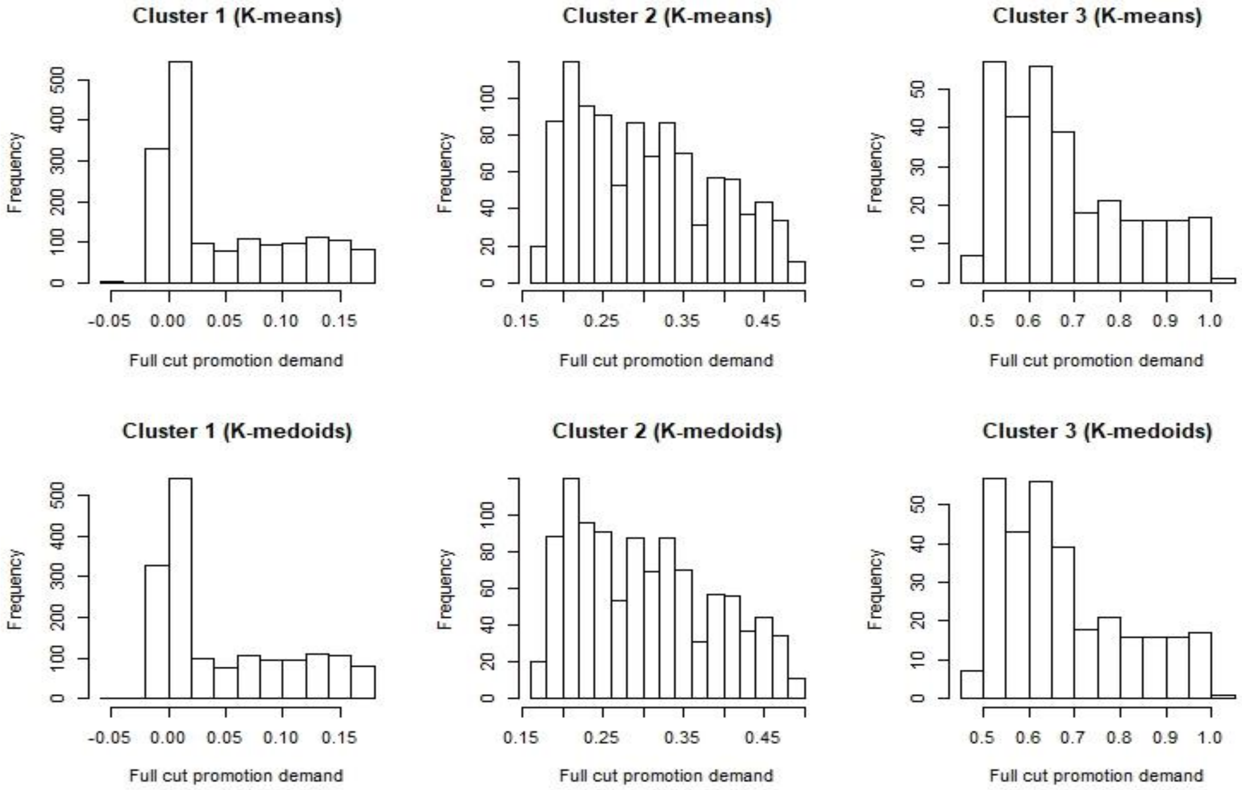


Figure 11. Distribution of each clusters K-means K-medoids

COMPARATIVE STUDY

Conditional Gradient as benchmark model to the proposed approach.

Conditional Gradient Model[1]

- Reformulate mixing distribution estimation as convex problem
- Optimize only through choice probabilities
- Via Conditional Gradient
- Algorithm and Framework provided in Appendix

Why do we choose this approach?

- Uses product-level features in aggregated data
- Estimate demand via choice probabilities and identify customer types
- New approach, recently submitted

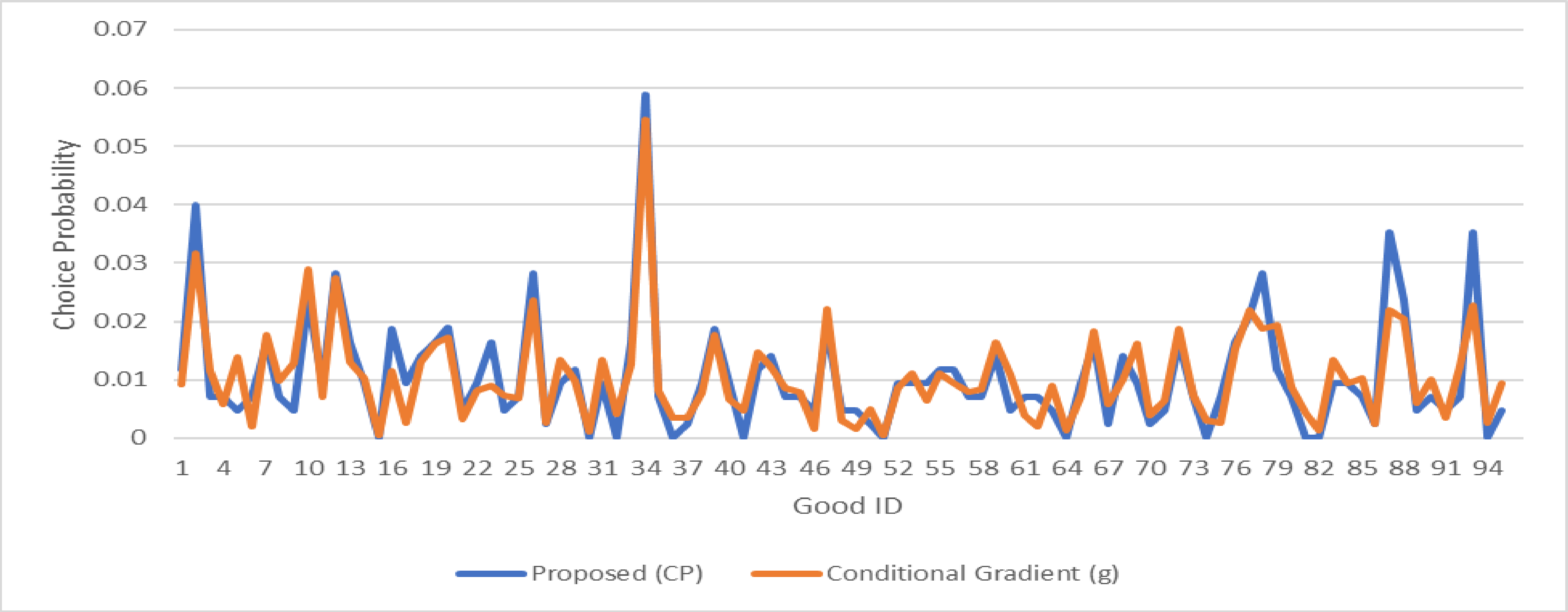
What are the challenges?

- Understanding the intuition of the framework
- Modifying objective function
- Translating algorithm into code

[1] Jagabathula, S., Subramanian, L., & Venkataraman, A. (2018). A Conditional Gradient Approach for Nonparametric Estimation of Mixing Distributions.

COMPARATIVE STUDY

Comparative study of customer types. Proposed vs Conditional Gradient.



$$g = \alpha^{(1)} * f^{(1)} + \alpha^{(2)} * f^{(2)} + \alpha^{(3)} * f^{(3)}$$

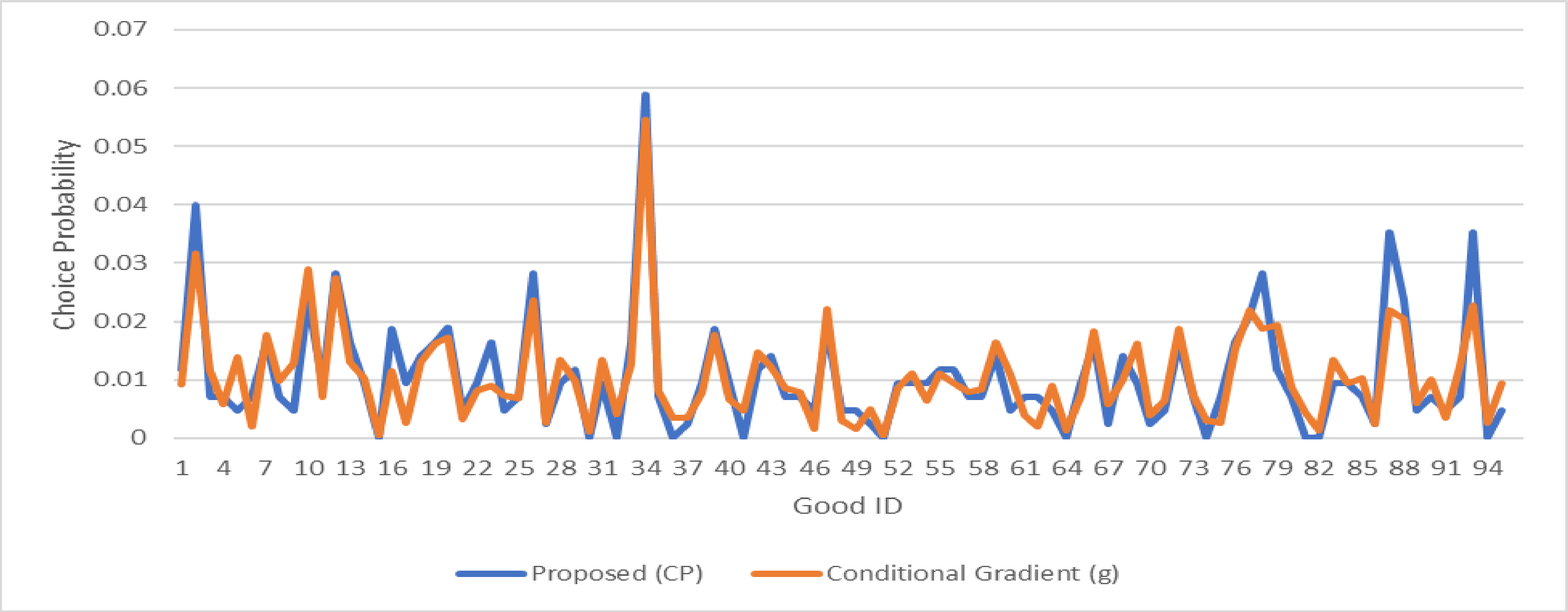
$$CP = w^1 * CP^1 + w^2 * CP^2 + w^3 * CP^3$$

CP^i is a choice probability vector across for cluster i in K-means

w^i is weight of observations for cluster i in K-means

COMPARATIVE STUDY

Comparative study of customer types. Proposed vs Conditional Gradient.



$percentdiff_j = |\boldsymbol{CP_j} - \boldsymbol{g_j}| / \boldsymbol{g_j}$ $\quad w_j = N_j / \sum_{j=1}^{95} N_j$

70.4% similarity in choice probability

$\sum_j w_j * percentdiff_j = 29.6\%$

Majority of error in choice probabilities occurs in good ID which are seldom purchased (lack of data)

Data-driven clustering models can yield a lower error with larger data

MACHINE LEARNING

Motivations for implementing machine learning models.

Investigate relationship between features and promotion driven demand

- Insights on how features (market price, number of goods) impact promotion driven demand
- Enables companies to fine-tune full-cut promotions beyond discount and threshold amount
- Understand how and why customers react towards the different discount & threshold amounts

MACHINE LEARNING

Data Processing, Features, Output.

Feature Categories	Features	Details	Examples
Quantitative	min_price	Price of the good_id which is the lowest amongst the other prices of good_id in an order.	¥82
	price-min_price	price minus the minimum price. order price of good_id that the user originally intends to purchase without full-cut promotion.	¥255
	cut_fav_amt	Discount amount in the order	¥0, ¥100, ¥200, ¥300
	market_price	Prices of product from competitors such as retail or e-commerce stores.	¥110
	total_goods_cnt	Total number of goods purchased in the orderorder.	1,2,8
Output (Target)	Promotion-driven demand	Apply logit transformation to price and then normalize to a range between 0 to 1. Target reflects the full-cut promotion driven-demand in each order.	1, 0.7478, 0.6818

METHODOLOGY

Regression Trees have the best model performance.

	MSE (Test set)	MAE (Test set)
Regression Tree (CART)	0.638%	4.223%
Regression Tree (Bagging)	0.582%	4.014%
Regression Tree (Boosted, GBM)	0.143%	1.178%
Regression Tree (XGBoost)	0.1491 %	1.392%
Regression Tree (H2O)	0.140%	1.226%
Regression Tree (Light GBM)	0.778%	5.051%
Artificial Neural Networks	0.847%	4.844%
K-Nearest Neighbors	1.283%	6.910%

METHODOLOGY

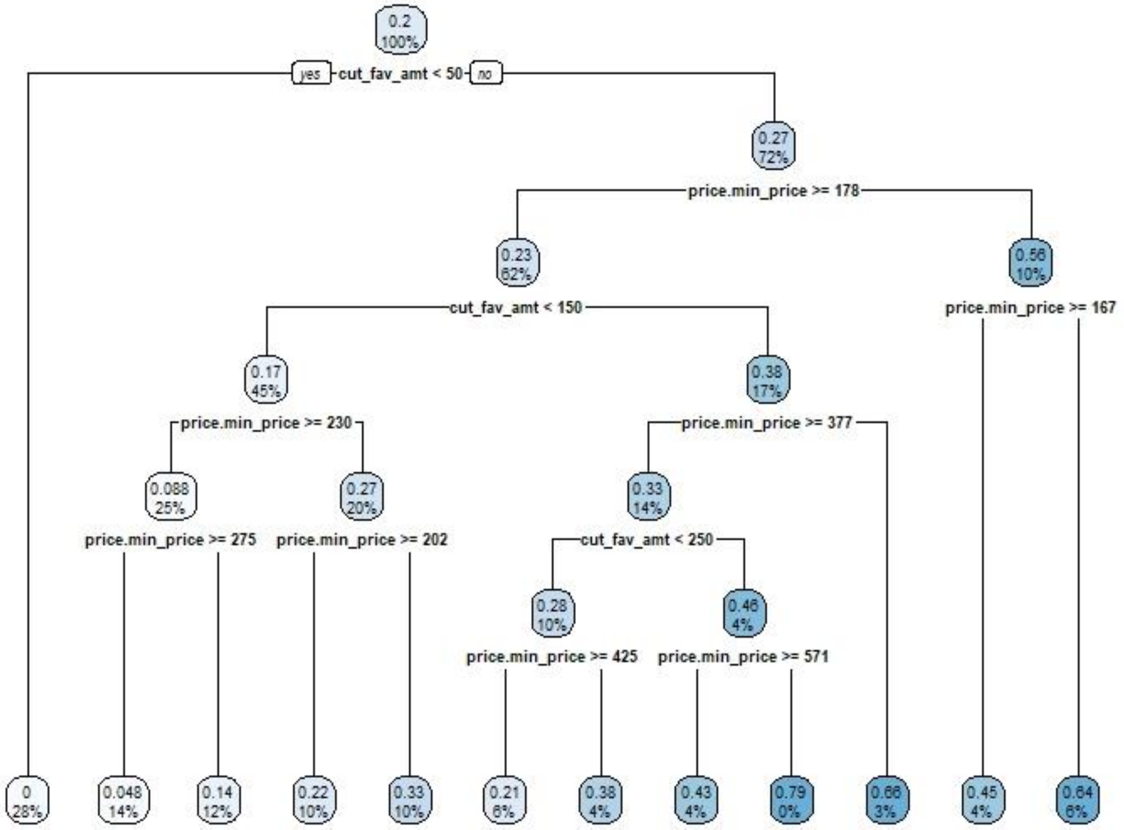


Figure 13 Regression Tree (CART) post parameter tuning

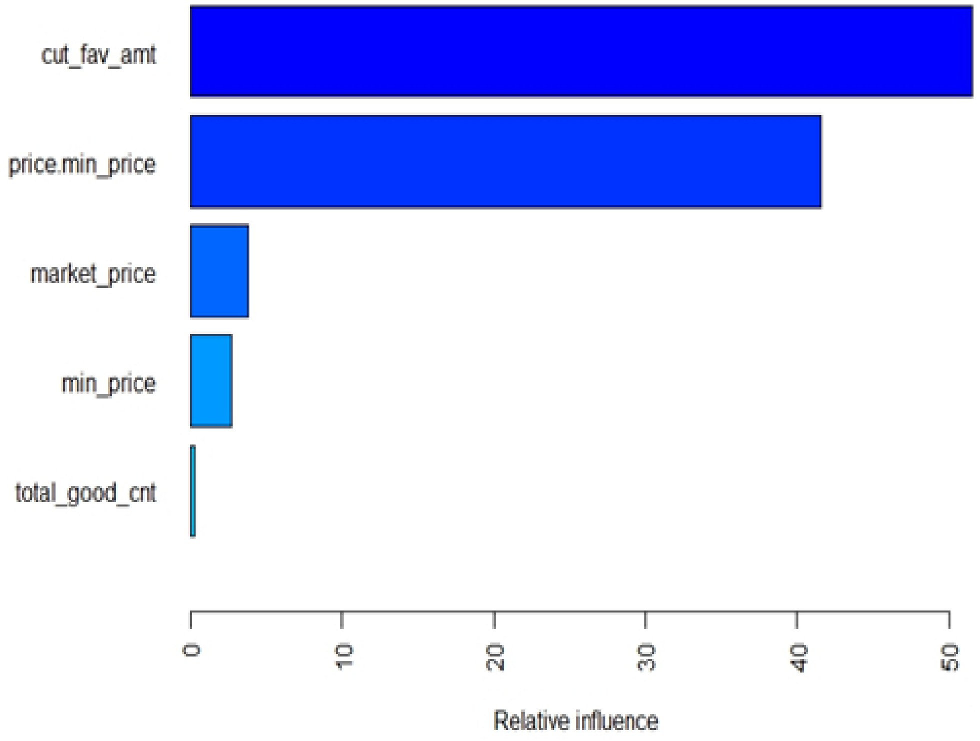


Figure 14. Variable Importance plot for Boosted trees, GBM

CONCLUSION

Proposed Methodology and Future Work.

1 Exploratory Data Analysis

2 Proposed Model for Promotion Sensitivity

3 Customer Segmentation & Comparative study

4 Machine Learning

5 Optimization + Future work

REFERENCES

[1] Jagabathula, S., Subramanian, L., & Venkataraman, A. (2018). A Conditional Gradient Approach for Nonparametric Estimation of Mixing Distributions.

[2] Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. Econometrica: Journal of the Econometric Society, 841-890.

THANK YOU!

APPENDIX

Example on Hierarchical Clustering.

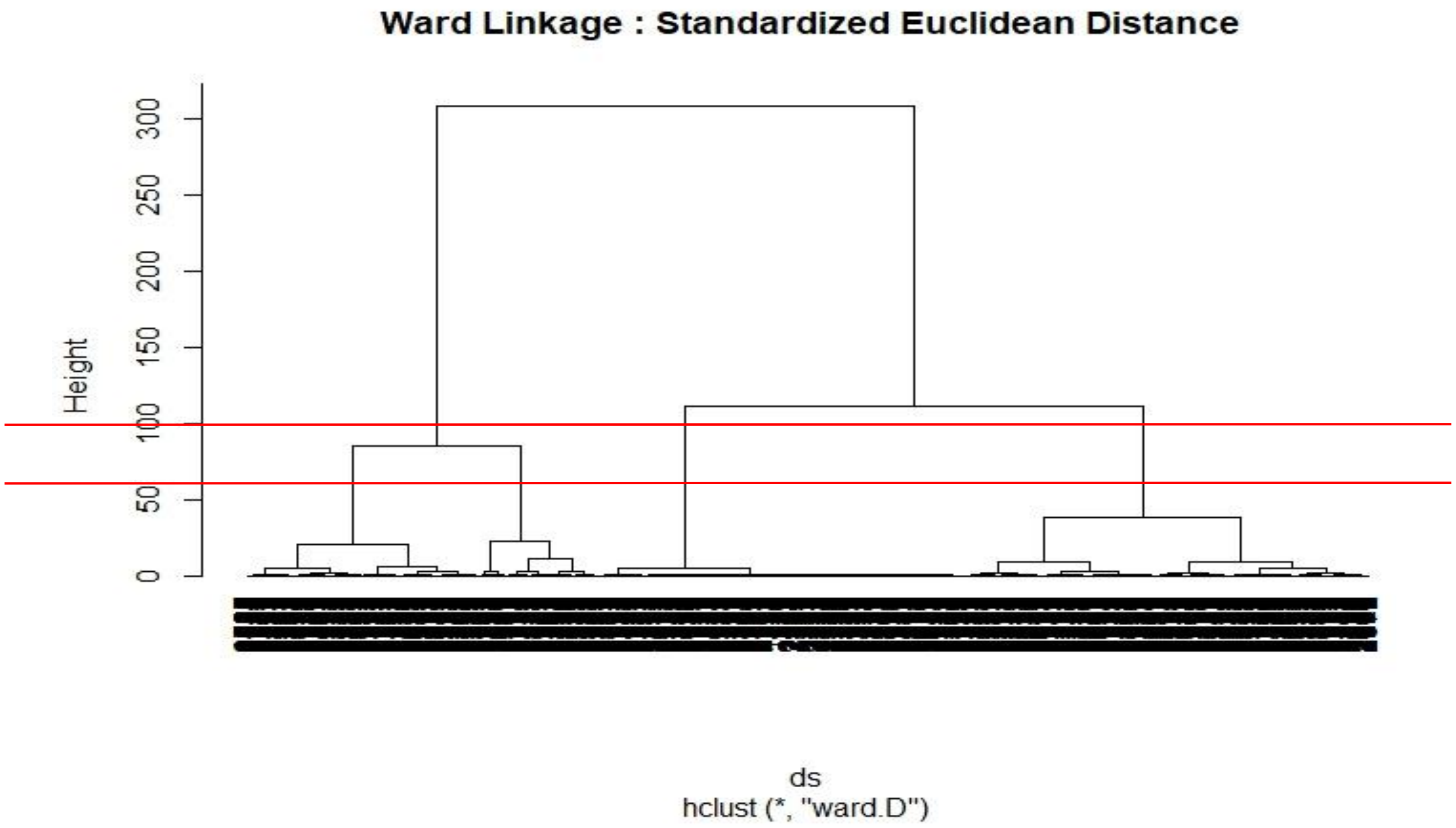


Figure 13. Dendrogram on Ward’s Linkage

CONDITIONAL GRADIENT SETUP

Goal: Identify customer types and choice probabilities

Estimate the best mixing distribution based on $\boldsymbol{Data}$ from all possible distributions Q

By minimizing the Squared Loss function,

$$\min_{Q} loss(\boldsymbol{g}(Q),\ \boldsymbol{Data})$$

$$loss(g(Q), \boldsymbol{Data}) = \frac{1}{2} \sum_{j \in [n]} \left(g_j(Q) - b_j\right)^2$$

$b_j = N_j / N$

$g_j(Q)$ represents the customer's choice probability on good j

Assuming that choice probability follows a Multinomial Logit model (MNL),

$$f_j(w) = \frac{\exp(w' \cdot z_j)}{\sum \exp(w' \cdot z_l)}$$

$$g_j(Q) = \int f_j(w) dQ(w)$$

CONDITIONAL GRADIENT SETUP

Instead of optimizing through all possible mixture distributions,

We can simply optimize through the convex hull of choice probability vectors:

$$\min_{Q} loss(\boldsymbol{g}(Q), \boldsymbol{Data}) \equiv \min_{g \in conv(\bar{P})} loss(\boldsymbol{g}, \boldsymbol{Data})$$

$$where$$

$$conv(\bar{P}) = \{\sum_{f \in F} \alpha_f \boldsymbol{f} : F \subset \bar{P} \text{ is finite and } \sum_{f \in F} \alpha_f = 1, \alpha_f \geq 0, \text{for all } \boldsymbol{f} \in F \}$$

$$P = \{\boldsymbol{f}(\boldsymbol{w}) : \boldsymbol{w} \in \mathbb{R}^D\} \qquad \boldsymbol{f}(\boldsymbol{w}) = \left(f_j(\boldsymbol{w}) : j \in [n]\right)$$

CONDITIONAL GRADIENT ALGORITHM

Algorithm

1: Initialize: $k = 0$; $g^{(0)}$ sampled based on standard uniform distribution, $\alpha^{(0)} = (1)$

2: While stopping condition is not met **DO**

3: $k \leftarrow k + 1$

4: Compute $\boldsymbol{f}^{(\boldsymbol{k})} \in \operatorname{\mathbf{argmin}}_{\boldsymbol{v} \in \overline{P}} \langle \boldsymbol{\nabla loss}(\boldsymbol{g}^{(k-1)}), \boldsymbol{v} - \boldsymbol{g}^{(k-1)} \rangle$ (Support finding step)

5: Compute $\boldsymbol{\alpha}^{(\boldsymbol{k})} \in \operatorname{\mathbf{argmin}}_{\boldsymbol{\alpha} \in \boldsymbol{\Delta}^{k}} \boldsymbol{loss}(\boldsymbol{\alpha_0 g}^{(0)} + \sum_{\boldsymbol{s=1}}^{\boldsymbol{k}} \boldsymbol{\alpha_s f}^{(s)})$ (Proportions update step)

6: Update $\boldsymbol{g}^{(\boldsymbol{k})} = \boldsymbol{\alpha}_{\boldsymbol{0}}^{(\boldsymbol{k})} \boldsymbol{g}^{(0)} + \sum_{\boldsymbol{s=1}}^{\boldsymbol{k}} \boldsymbol{\alpha}_{\boldsymbol{s}}^{(\boldsymbol{k})} \boldsymbol{f}^{(s)}$

7: **end while**

8: Output: Probabilities $\boldsymbol{\alpha}_{\boldsymbol{1}}^{(\boldsymbol{3})} = 0.002, \boldsymbol{\alpha}_{\boldsymbol{2}}^{(\boldsymbol{3})} = 0.996, \boldsymbol{\alpha}_{\boldsymbol{3}}^{(\boldsymbol{3})} = 0.002$ for each corresponding customer type $\boldsymbol{f}^{(\boldsymbol{1})}, \boldsymbol{f}^{(\boldsymbol{2})}, \boldsymbol{f}^{(\boldsymbol{3})}$.

$\boldsymbol{f}^{(\boldsymbol{k})}$ is a choice probability vector across all 95 types of products

$\boldsymbol{\alpha}^{(\boldsymbol{k})}$ is the likelihood that a customer′ s choice probability follows $\boldsymbol{f}^{(\boldsymbol{k})}$

$\boldsymbol{g} = \boldsymbol{\alpha}_{\boldsymbol{1}}^{(\boldsymbol{3})} * \boldsymbol{f}^{(1)} + \boldsymbol{\alpha}_{\boldsymbol{2}}^{(\boldsymbol{3})} * \boldsymbol{f}^{(2)} + \boldsymbol{\alpha}_{\boldsymbol{3}}^{(\boldsymbol{3})} * \boldsymbol{f}^{(3)}$ is the average customer’s choice probability

OPTIMIZATION

Goal: Obtain threshold & discount amount that maximizes revenue

By identifying the choice probability for each type of promotion,

Assuming customers seek to maximize utility in their purchases,

$$\max_{\boldsymbol{x}\in\Delta_n} \mathrm{E}\left[\sum_{j=0}^{n}\left(U_{ij}-\alpha p_j\right)x_{kj}+\gamma_i\cdot\mathbf{1}_{\{Order\ amount\geq\theta\}}\cdot\delta\right]+\varepsilon$$

$where$

$\Delta_n=\{\boldsymbol{x}\in\mathbb{R}_{+}^{n+1}|\sum_{j=0}^{n}x_{ij}=1\}$ constraint space for choice probabilities across goods

$j \in product\ universe\ [n]$ and $j=0$ denotes the outside option

x_{kj} denotes the choice probability, p_j denotes the price of good j

α denotes price sensitivity

OPTIMIZATION

Goal: Obtain threshold & discount amount that maximizes revenue

By identifying the choice probability for each type of promotion,

Assuming customers seek to maximize utility in their purchases,

$$\max_{\boldsymbol{x} \in \Delta_n} \mathrm{E}\left[\sum_{j=0}^{n} (U_{ij} - \alpha p_j) x_{kj} + \gamma_i \cdot \mathbf{1}_{\{Order\ amount \geq \theta\}} \cdot \delta\right] + \varepsilon$$

$where$

$\gamma_i \cdot \mathbf{1}_{\{Order\ amount \geq \theta\}} \cdot \delta$ positive utility from full-cut promotion

γ_i denotes the customer i's sensitivity parameter to full-cut promotion

δ denotes the discount amount

θ denotes the threshold amount

OPTIMIZATION

Goal: Obtain threshold & discount amount that maximizes revenue

By identifying the choice probability for each type of promotion,

Assuming customers seek to maximize utility in their purchases,

$$\max_{\boldsymbol{x} \in \Delta_n} \mathrm{E}\left[\sum_{j=0}^{n} \left(U_{ij} - \alpha p_j\right) x_{kj} + \gamma_i \cdot \mathbf{1}_{\{Order\ amount \geq \theta\}} \cdot \delta\right] + \varepsilon$$

$where$

$U_{ij} = w_k' * z_{ij}$ denotes the customer's utility for good j

w_k' denotes the customer's preferences

z_{ij} denotes the good's features

OPTIMIZATION

Goal: Obtain threshold & discount amount that maximizes revenue.

w_k' is an input for customer's utility model to optimize full-cut promotion

To obtain the preference vector w_k' for each customer type k,

Minimize the error between customer's utility and choice probability,

$$\min_{\boldsymbol{w_k}} \sum_{j=1}^{n} \boldsymbol{w_k'} \cdot z_j \underbrace{-p_j}_{\text{Constant}} \underbrace{- \ln(x_{kj}) + \ln(x_{k0})}_{\text{Based on clustering of promotion-driven demand } y}$$

where

x_{kj} denotes the choice probability of customer type k, product j

x_{k0} denotes the choice probability of customer type k, not buying product j

w_k' denotes the customer type k preferences

z_j denotes the good j features

OPTIMIZATION

Optimization to obtain threshold & discount amount.

Goal: To identify threshold & discount amount that maximizes total revenue across all products

For each (θ, δ), we obtain the $\boldsymbol{x}$

With the profit margin vector of all goods, $\boldsymbol{p}$,

$$\max_{(\theta,\delta)\in(\boldsymbol{\theta},\boldsymbol{\delta})} \boldsymbol{x} * \boldsymbol{p}$$

(θ^*, δ^*) with a highest revenue, $(\boldsymbol{x} \cdot \boldsymbol{p})^*$ will be the optimal threshold and discount amount

Comparative study with Discrete Choice models

$$\ln\left(\frac{y_i}{1-y_i}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\beta_0 = -666.6 \qquad \beta_1 = 3.38, \qquad \beta_2 = 3.35$$

x_1: minimum price of the good in an order

x_2: order amount discounted with the minimum priced good

	Misclassification rate
Binary Logit Model	6.487 %
XGBoost	1.669%