Towards User-in-the-Loop Online Fashion Size Recommendation with Low Cognitive Load

Leonidas Lefakis, Evgenii Koriagin
Julia Lasserre, Reza Shirvany

Zalando SE

Size Recommendation

What are we trying to solve?

Given a customer c and an article a recommend size s^*

$$s^* = \operatorname{argmax}\left(P(R = kept, S = s | C = c, A = a)\right)$$

where R is the return status of the order.

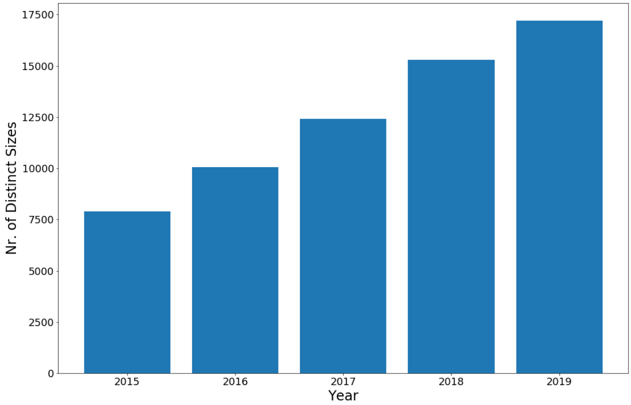
Why is it important?

- Customer Experience.
- Platform Profits.
- Environmental Impact.

Why is it difficult?

- Limitations and Coarseness of Sizing Systems.
- Grading from Target Sizes.
- Variety of Sizing Systems.
- Non-Standardization of Sizing Systems.
- Vanity Sizing.
- Subjective Aspect of Size and Fit.

Size Problem Complexity



Distinct Apparel Sizes

Assume customer c has bought article a in multiple sizes returning some.

- Customer offset

$$o_{ca} = \frac{1}{K_{ca}} \sum_{k=1}^{K_{ca}} s_{cak} - \frac{1}{R_{ca}} \sum_{r=1}^{R_{ca}} s_{car}$$

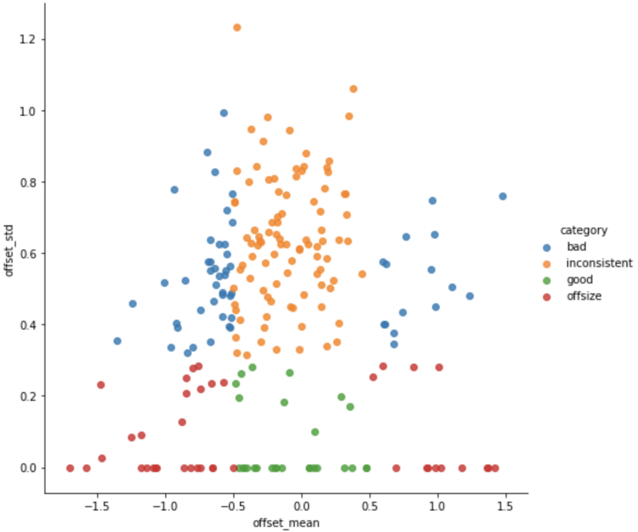
- Article offset distribution

$$\mathcal{N}_a \left(\frac{1}{N_C} \sum_{c=1}^{N_C} o_{ca}, \frac{1}{N_C - 1} \sum_{c=1}^{N_C} (o_{ca} - \mu_a)^2 \right)$$

If article a_i belonging to brand b has sales w_i.

- Brand offset distribution

$$\mathcal{N}_b\left(\frac{1}{\sum_{i=1}^{N_A} w_i}\sum_{i=1}^{N_A} w_i \mu_{a_i}, \frac{1}{\sum_{i=1}^{N_A} w_i}\sum_{i=1}^{N_A} w_i (\mu_{ai} - \mu_b)^2\right)$$

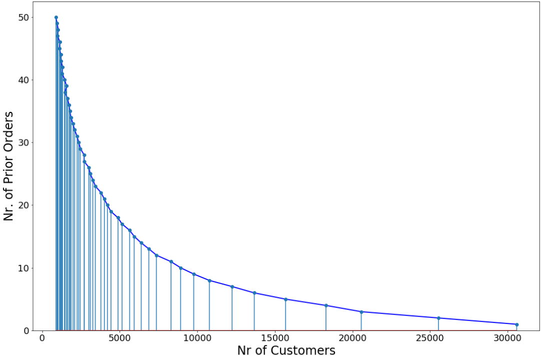


Brand Size Offsets σ,μ

- Size tables, article measurements.
- Article based size advice.
- Personalized order history based size reco.
- Customer in the loop personalized size reco.
- Computer vision and 3D approaches.

- Size tables, article measurements.
- Article based size advice.
- Personalized order history based size reco.
- **Customer in the loop personalized size reco**.
- Computer vision and 3D approaches.

Cold Start



Number of customers vs. number of prior orders.

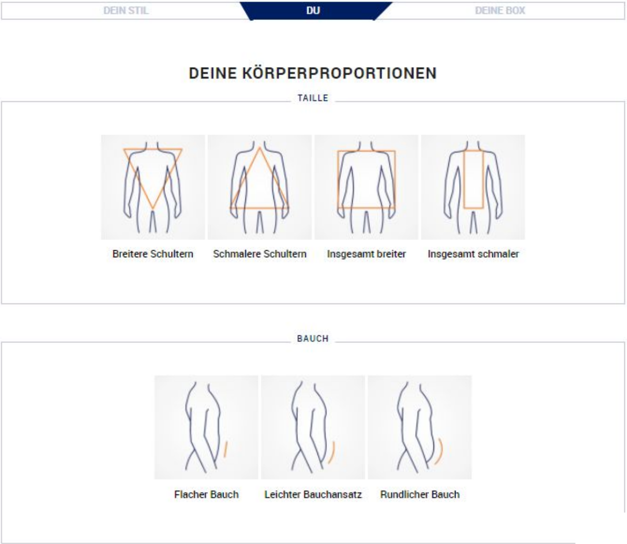
We want to predict customer size in the female upper garment category without the benefit of prior order information.

We want to predict customer size in the female upper garment category without the benefit of prior order information.

→ **Zalon** (styling service)

Type	Features
Overall	weight, height, age, gender
Upper body	top size, shirt collar size, shirt fit, proportion belly, top fit, proportion shoulder-waist, bust number, bust cup size, proportion shoulder-hip, blazer size
Lower body	pants size, jeans length, jeans width, proportion waist, pants waist-height , shoe size

Questionnaire Example



Type	Features
Overall	weight, height, age, gender
Upper body	**top size**, shirt collar size, shirt fit, proportion belly, top fit, proportion shoulder-waist, bust number, bust cup size, proportion shoulder-hip, blazer size
Lower body	pants size, jeans length, jeans width, proportion waist, pants waist-height , shoe size

OK so problem solved!

Customers report their sizes

OK so problem solved!

Customers report their sizes $\rightarrow$ we recommend them their sizes!

OK so problem solved!

Customers report their sizes $\rightarrow$ we recommend them their sizes!

Unfortunately that is not how it works :(

OK so problem solved!

Customers report their sizes

- Customers only buy their stated size $\sim 50\%$ of the time.

OK so problem solved!

Customers report their sizes

- Customers only buy their stated size $\sim 50\%$ of the time.
- Customers tend to under-report their size.

OK so problem solved!

Customers report their sizes

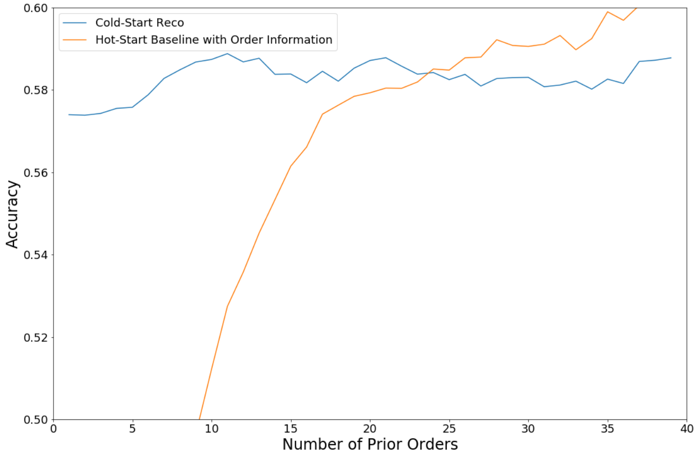
- Customers only buy their stated size $\sim 50\%$ of the time.
- Customers tend to under-report their size.
- Male Customers are more likely to under-report their size.

Then let's do ML!

Cross-validation experiments highlighted the benefits of **Gradient Boosted Trees.**

- An ensemble of Classification Trees built sequentially with boosting.
- Strong Performance + Interpretability + Robust to Overfitting.

Cold-Start Performance



The presented solution requires us to ask 20 size-related questions from the customer.

$\rightarrow$ This requires high engagement from the customer!

Cognitive Load

The presented solution requires us to ask 20 size-related questions from the customer.

→ This requires high engagement from the customer!

Can we retain performance while reducing cognitive load?

Performance vs Cognitive Load

- Using Weight, Height, and Top Size performs very well.
- Top Size + Brand Information is a close second.

- Using Weight, Height, and Top Size performs very well.
- Top Size + Brand Information is a close second.

Top Size + Brand are not personal data!

Customer prefer dislike giving weight and height.

- Using Weight, Height, and Top Size performs very well.
- Top Size + Brand Information is a close second.

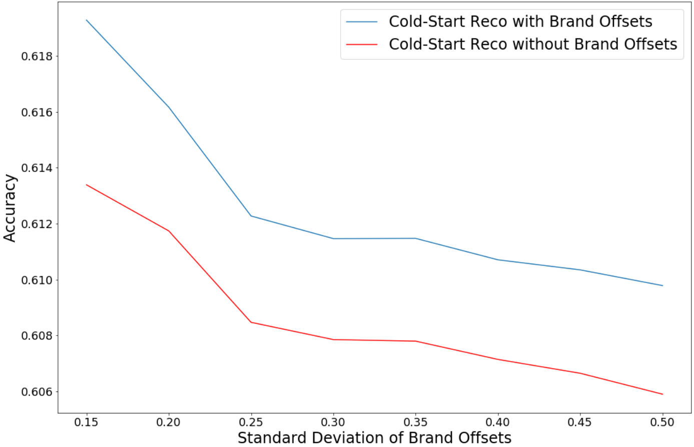
Top Size + Brand are not personal data!

Customer prefer dislike giving weight and height.

Top Size + Brand is the best alternative solution.

Dialogue Tile

Impact of Brand Offsets



Can we also leverage prior orders?

None of the presented solutions use order information.

Can we also leverage prior orders?

None of the presented solutions use order information.

→ Can we build a personalized reco that leverages both forms of data?

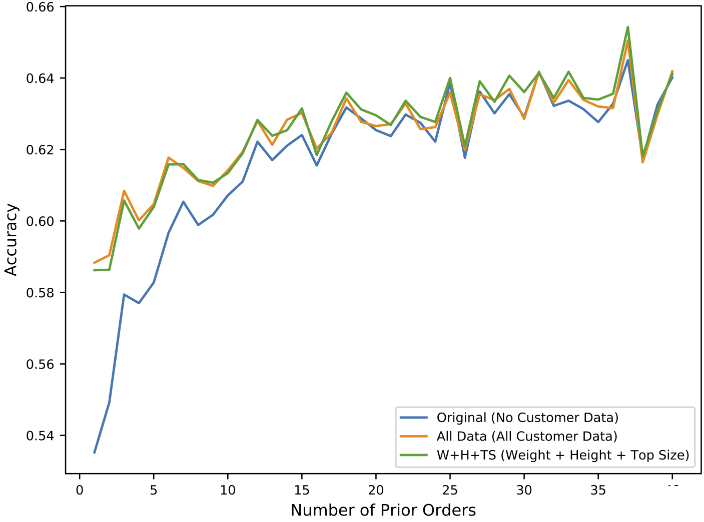
Can we also leverage prior orders?

None of the presented solutions use order information.

→ Can we build a personalized reco that leverages both forms of data?

→ Do our findings continue to hold in this case?

MetalSF (Lasserre et al. 2020)



Questions?