Size Normalization
Scale of Size Normalization Problem

- 83,977 distinct sizes
- 492,567 distinct size runs
Notes

1. **Normalization on Category-Brand-Size**
   
   eg. \((\text{Men's Shoes, Nike, "12C")} \rightarrow 54\)

2. **Only using transaction data**
   
   No Feedback, eg. "Too small" or "Too big"
   
   No extra user or item information, eg. height, weight, etc.
Proposed Solution

$$\min \sum_{i=0}^{\vert B \vert} \sum_{j=i+1}^{\vert B \vert} \sum_{m=0}^{\vert S_{b_i} \vert} \sum_{n=0}^{\vert S_{b_j} \vert} F(b_i,s_m),(b_j,s_n) \cdot (x_{b_i,s_m} - x_{b_j,s_n})^2$$
Proposed Solution

\[
\min \sum_{i=0}^{\mid B \mid} \sum_{j=i+1}^{\mid B \mid} \sum_{m=0}^{\mid S_{b_i} \mid} \sum_{n=0}^{\mid S_{b_j} \mid} F(b_{i,s_m}, (b_{j,s_n}) \times (x_{b_{i,s_m}} - x_{b_{j,s_n}})^2
\]

For every brand
Proposed Solution

For every brand
For every size in the brand

\[
\min \sum_{i=0}^{\left| \mathcal{B} \right|} \sum_{j=i+1}^{\left| \mathcal{B} \right|} \sum_{m=0}^{\left| S_{b_i} \right|} \sum_{n=0}^{\left| S_{b_j} \right|} F(b_i, s_m, (b_j, s_n)) \cdot (x_{b_i, s_m} - x_{b_j, s_n})^2
\]
Proposed Solution

For every brand
For every size in the brand

\[
\min \sum_{i=0}^{\left| \mathcal{B} \right|} \sum_{j=i+1}^{\left| \mathcal{B} \right|} \sum_{m=0}^{\left| S_{b_i} \right|} \sum_{n=0}^{\left| S_{b_j} \right|} F(b_i, s_m, b_j, s_n) \cdot (x_{b_i, s_m} - x_{b_j, s_n})^2
\]

Minimize distance in shared space.
Proposed Solution

For every brand in the brand

For every size

Weighted by the copurchase frequency.

Minimize distance in shared space.

\[
\min \sum_{i=0}^{\left| \mathcal{B} \right|} \sum_{j=i+1}^{\left| \mathcal{B} \right|} \sum_{m=0}^{\left| S_{b_i} \right|} \sum_{n=0}^{\left| S_{b_j} \right|} F(b_i, s_m), (b_j, s_n) \cdot (x_{b_i, s_m} - x_{b_j, s_n})^2
\]
Proposed Solution

\[
\begin{align*}
\text{min} & \quad \sum_{i=0}^{\lvert \mathcal{B} \rvert} \sum_{j=i+1}^{\lvert \mathcal{B} \rvert} \sum_{m=0}^{\lvert S_{b_i} \rvert} \sum_{n=0}^{\lvert S_{b_j} \rvert} F(b_i,s_m),(b_j,s_n) \cdot (x_{b_i,s_m} - x_{b_j,s_n})^2 \\
+ & \quad \sum_{i=0}^{\lvert \mathcal{B} \rvert} \frac{0.1}{\lvert S_{b_i} \rvert} \left( x_{b_i,s_{\lvert S_{b_i} \rvert}} - x_{b_i,s_0} \right) \\
\text{s.t.} & \quad x_{b_i,s_{m+1}} - x_{b_i,s_m} \geq 0.1 \quad \forall b_i \in \mathcal{B}, m \in S_{b_i}
\end{align*}
\]

For every brand For every size in the brand Weighted by the copurchase frequency. Minimize distance in shared space.

Make sure that sizes on the extremes gets assigned reasonable values.

eg. XXXS, XXXL
Results - Womens Shoes

Size Type

Normalized Size
Results - Womens Dresses
## Results

<table>
<thead>
<tr>
<th></th>
<th>First Year Accuracy (Training Set)</th>
<th>Second Year Accuracy (Test Set)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GD</td>
<td>QP</td>
</tr>
<tr>
<td>Women’s shoes</td>
<td>62%</td>
<td>62%</td>
</tr>
<tr>
<td>Women’s dresses</td>
<td>58%</td>
<td>58%</td>
</tr>
</tbody>
</table>

Train: Off by **1-2%**  
Test: Off by **up to 8%**
Automated Fashion Size Normalization

Eddie S.J. Du, Chang Liu, David H. Wayne