Session-based Complementary Fashion Recommendations

Jui-Chieh Wu, José Antonio Sánchez Rodríguez, Humberto Corona
WE OFFER A SUCCESSFUL AND CURATED ASSORTMENT

> 400,000 articles from
> 2,000 international brands

LOCALIZATION of the assortment

HIGHLY EXPERIENCED category management

11 private labels

CURATED SHOPPING with Zalando
About Zalando

Assortment

> 400,000

Brands

> 2,000

Active Customers

> 28 m
What is Complementary Item Recommendation?
Our Baseline

- Based on item-item collaborative filtering
- High score items in different category
- Items similar to high score items if not enough recommendations
Limitations of the Baseline

- Static recommendation for everyone
- Low CTR
- Low conversion rate
Problem Statement

For a given user with an interaction history $x_h$ and the anchor item $x_t$, select a list of complementary items $y_1, y_2, \ldots, y_k$ from a set of candidates $y$.

$$y_1, y_2, \ldots, y_k \sim p(y \mid x_h, x_t)$$

How We Define Complementary Relationship

Two items $x_i$ and $x_j$ are complementary if they

1. Belong to different categories (shoe v.s. trousers)
2. Belong to two fashion-compatible categories
Problem Statement

\[ y_1, y_2, \cdots, y_k \sim p(y | x_h, x_t) \]

- Complementary of \( x_t \)
- Whole Catalog

- Learn from the existing user response on the current baseline
- Learn from the re-sampled dataset
Creating a More Representative Dataset

Training a new model on top of the training data coming from the baseline constraints the capacity of the abstractions learned by the new model.

**Solution**
Instead of learning from the user behavior we observed on the current product, we sample behaviors from the *user interaction history* that satisfy our definition of complementary.
Creating a More Representative Dataset

Timeline

User Interaction history
Creating a More Representative Dataset

User Interaction History

Generated Sequences

Timeline

Anchor

Target

History

Target Window
Model Architecture

cross-entropy-loss

\[ p(y \mid x_h, x_i) \]

Softmax

Score [1 x O]

Trilinear Combination

- User History Encoder
  - User History [M]
- Current Context Encoder
  - Anchor Item [1]
- Candidate Encoder
  - Candidates [O]

Liu et al., STAMP: ShortTerm Attention/Memory Priority Model for Session-based Recommendation. (KDD 2018).
Our Adjustments - Add Long Term Signals

\[
p(y \mid x_h, x_t)
\]

Softmax

Score [1 x O]

cross-entropy-loss

Trilinear Combination

Order History Encoder

User History Encoder

Current Context Encoder

Candidate Encoder

Order History [M]

User History [M]

Anchor Item [1]

Candidates [O]
Our Adjustments - Additive Combination Function

\[ p(y | x_h, x_t) \]

Softmax

Score [1 x O]

\[ (h_o + h_s + h_t)^T x_{ci} \]
Our Adjustments - Context Information Added

STAMP

v.s.

Our Model

Item Embedding

Image Embedding

Category Embedding

Time Difference Embedding

FFN

1 + Y

Concatenation
## Evaluation Results - Offline

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall@5</th>
<th>Order Recall@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Method</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>Collaborative Filtering</td>
<td>0.29</td>
<td>0.24</td>
</tr>
</tbody>
</table>
## Evaluation Results - Online A/B Test

<table>
<thead>
<tr>
<th></th>
<th>CTR</th>
<th># Items Ordered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Method</td>
<td>+6.23%</td>
<td>+3.24%</td>
</tr>
<tr>
<td>Model</td>
<td>Recall@5</td>
<td>Order Recall@5</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td>STAMP</td>
<td>0.221</td>
<td>0.206</td>
</tr>
<tr>
<td>STAMP + Long Term Signal</td>
<td>0.241</td>
<td>0.223</td>
</tr>
<tr>
<td>STAMP + Context Information</td>
<td>0.258</td>
<td>0.255</td>
</tr>
<tr>
<td>STAMP + Image Feature</td>
<td>0.264</td>
<td>0.240</td>
</tr>
<tr>
<td><strong>Our Method</strong></td>
<td><strong>0.264</strong></td>
<td><strong>0.267</strong></td>
</tr>
</tbody>
</table>
Conclusion

- We devised a personalized complementary fashion recommender that outperformed the baseline in an A/B test.

- We tailored STAMP, one of the state-of-the-art session recommenders, and yields better performance on our dataset.

- Through the ablation test, we assures the efficacy of the model improvements.
QUESTIONS?