Basket-Sensitive Personalized Item Recommendation

Duc-Trong Le, Hady W. Lauw, Yuan Fang

IJCAI 2017
Melbourne, Australia
Outline

• **Motivating examples and Models**
  • Factorization Machine (FM)
  • Basket-Sensitive Factorization Machine (BFM)
  • Constrained Basket-Sensitive Factorization Machine (CBFM)

• **Experiments** on BeiRen, Foursquare
Personalized Item Recommendation

Learn a real-value function from adoptions to rank candidate items

\[ F(u_i, v_j; \Theta) \]

Matrix Factorization:

\[ F(u_i, v_j; \Theta) \propto \phi_i \ast \phi_j \text{ where } \phi_i, \phi_j \in \mathbb{R}^K \]
Factorization Machine (FM)

For each adoption \(<u_i, v_j>\):

**User** \(u_i\)  \hspace{1cm}  **Adopted Item** \(v_j\)  \hspace{1cm}  **Adoption** \(\delta\)

\[ h = \begin{array}{ccccccc}
0 & \ldots & 1 & \ldots & 0 & 0 & 0 & \ldots & 1 & \ldots & 0 & \ldots & 0 & 0 & 1
\end{array} \]

**Adoption Estimation**  \hspace{1cm}  **Global & Item Bias**  \hspace{1cm}  **Interactive Features**

\[
F(h; \Theta) = \mu_0 + \sum_{i=1}^{p} \mu_i h_i + \sum_{i=1}^{p} \sum_{j=i+1}^{p} h_i h_j \sum_{f}^{K} \phi_{i,f} \phi_{j,f}
\]

**Optimization:**

\[
OPT_{FM}(T) = \arg\min_{\Theta} \sum_{\{h\}} l(F(h; \Theta), \delta)
\]

Loss function
The Notion of Basket – Shopping Scenarios

Shopping cart on …

Monday

Basket
Sushi

Mary

Thursday

Basket
Chili Crab

Saturday

Basket
Cake

Items in each basket might be correlated
Basket-Sensitive Personalized Item Recommendation

Given the current basket, what items should Mary add?

Current Basket

Basket Items

Learn a real-value function to rank candidate items

\[ F(u_i, B_i, v_j; \Theta) \]
Modeling Association Types - $\gamma_1$

User - Candidate - Item

Mary likes Milk

Capture personalization
Modeling Association Types - $\gamma_2$

Milk & Baking Items

Capture correlations

Candidate Item

Basket Items
Modeling Association Types - $\gamma_3$

Among Basket Items

Capture correlations in the current basket
Modeling Association Types - $\gamma_4$

Mary likes Baking Material

Capture personalization (Maybe redundant)
Real-Valued Function with Association Types

\[ F(u_i, B_i, v_j; \Theta) \propto \gamma_1 x_i^T y_j + \gamma_2 \sum_{v_k \in B_i} y_j^T z_k \]

Adoption Estimation
User & Candidate Item
Candidate Item & Basket Items

+ \gamma_3 \cdot \sum_{(v_k \neq v_{k'}) \in B_i} z_{k'}^T z_k' + \gamma_4 \cdot \sum_{v_k \in B_i} x_i^T z_k

Among Basket Items
User & Basket Items

where \( x_i, y_j, z_k \in \mathbb{R}^K \) are latent vectors; \( \gamma_1, \gamma_2, \gamma_3, \gamma_4 \in \{0,1\} \)
Basket-Sensitive Factorization Machine (BFM)

Given N users, M items;
Tuple \( t = < u_i, B_i, v_j, \delta > \in T; t. \delta \in \{-1, 1\}; \)

\[ F(h; \Theta) = \mu_0 + \sum_{i=1}^{p} \mu_i h_i + \gamma_1 \sum_{i=1}^{N} \sum_{j=N+1}^{N+M} h_i h_j (\phi_i^T \phi_j) + \gamma_2 \sum_{i=N+1}^{N+M} \sum_{j=N+M+1}^{p} h_i h_j (\phi_i^T \phi_j) \]

\[ + \gamma_3 \sum_{i=N+M+1}^{p} \sum_{j=i+1}^{p} h_i h_j (\phi_i^T \phi_j) + \gamma_4 \sum_{i=1}^{N} \sum_{j=N+M+1}^{p} h_i h_j (\phi_i^T \phi_j) \]
Same-Intent Tuples

$t_1 = \langle \text{Mary}, \{A, C, B\}, D, 1 \rangle$

$t_2 = \langle \text{Mary}, \{A, B, D\}, C, 1 \rangle$

Same-Intent Tuples
Constrained BFM (CBFM)

• Given same-intent tuples \((t_1, t_2)\), we expect

\[
PMI(t_1, t_2) \times \left( \mathcal{F}(h^{t_1}; \Theta) - \mathcal{F}(h^{t_2}; \Theta) \right)^2
\]

should be small

Pointwise Mutual Information  Adoption Estimation Difference

• Not all tuple pairs have strong intended effect, only consider the constraint in optimization task:

\[
PMI(t, t^m) \times \left( \mathcal{F}(h^t; \Theta) - \mathcal{F}(h^{t^m}; \Theta) \right)^2
\]

where \(t, t^m\) are the current tuple and its same-intent tuple that has the maximum \(\mathcal{F}(h^{t^m}; \Theta)\) score
Optimization

\[
\text{OPT\_BFM}(T) = \arg\min_{\Theta} \left[ \sum_{t \in T} - \ln(\sigma(\mathcal{F}(h^t; \Theta) \times t.\delta)) + \sum_{\theta \in \Theta} \lambda_\theta \theta^2 \right]
\]

\[
\text{OPT\_CBFM}(T) = \arg\min_{\Theta} \left[ \sum_{\theta \in \Theta} \lambda_\theta \theta^2 + \sum_{t \in T} \left\{ - \ln(\sigma(\mathcal{F}(h^t; \Theta) \times t.\delta)) \right\} + \frac{\alpha}{2} \times PMI(t, t^m) \times \left( \mathcal{F}(h^t; \Theta) - \mathcal{F}(h^{tm}; \Theta) \right)^2 \right]
\]

Stochastic Gradient Descent with Adaptive Learning Rate
\[
\alpha, \lambda_\theta \in \mathbb{R}^+; \quad \sigma(a) = 1/(1 + e^{-a})
\]
Outline

• Motivating examples and Models
  • Factorization Machine (FM)
  • Basket-Sensitive Factorization Machine (BFM)
  • Constrained Basket-Sensitive Factorization Machine (CBFM)

• Experiments on BeiRen, Foursquare
Experimental Setup

• **Task**: Next adoption recommendation
  – For each testing tuple \( t = < u_i, B_i, v_j, \delta > \)
  – Hide the observed adoption \( v_j \), require each model to product a ranked list of candidates for \( u_i \) based on \( B_i \)

• **Datasets**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#User</th>
<th>#Item</th>
<th>#Transaction</th>
<th>Avg. #Item / Transaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>BeiRen (Grocery shopping)</td>
<td>9245</td>
<td>5581</td>
<td>87224</td>
<td>6.1</td>
</tr>
<tr>
<td>Foursquare (Point-of-Interest)</td>
<td>1548</td>
<td>3619</td>
<td>31377</td>
<td>2.7</td>
</tr>
</tbody>
</table>

*Pre-processing*: Filter out too few or too popular items; Sample negative tuples \( (t. \delta = -1) \)

• **Metric**: *Half-life Utility (HLU)*

\[
HLU = \frac{1}{|T_{test}|} \times C \times \sum_{t \in T_{test}} 2^{1-r_t} ; C = 100, \beta = 5
\]
What are effects of the Association Types?

Higher is better

<table>
<thead>
<tr>
<th>Association</th>
<th>BeiRen</th>
<th>Foursquare</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$ $\gamma_2$ $\gamma_3$ $\gamma_4$</td>
<td>1 0 0 0</td>
<td>1.94</td>
</tr>
<tr>
<td>$\gamma_1$ $\gamma_2$ $\gamma_3$</td>
<td>1 1 0 0</td>
<td>3.35</td>
</tr>
<tr>
<td>$\gamma_1$ $\gamma_2$ $\gamma_3$</td>
<td>1 1 1 0</td>
<td>3.75*</td>
</tr>
<tr>
<td>$\gamma_1$ $\gamma_2$ $\gamma_3$</td>
<td>1 1 0 1</td>
<td>3.59</td>
</tr>
<tr>
<td>$\gamma_1$ $\gamma_2$ $\gamma_3$ $\gamma_4$</td>
<td>1 1 1 1</td>
<td>3.74</td>
</tr>
</tbody>
</table>

$\gamma_1$ User & Candidate Item
$\gamma_2$ Candidate & Basket Items
$\gamma_3$ Among Basket Items
$\gamma_4$ User & Basket Items

Latent dimension $K = 8$
How does the constraint influence the prediction performance?

Higher is better

Latent dimension $K = 8$;
Do CBFM & BFM outperform the Association Rules-based method?

Higher is better

Latent dimension $K = 8$

<table>
<thead>
<tr>
<th></th>
<th>CBFM</th>
<th>BFM</th>
<th>ASR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BeiRen</td>
<td>3.89</td>
<td>3.75</td>
<td>3.74</td>
</tr>
<tr>
<td>Foursquare</td>
<td>10.92</td>
<td>8.51</td>
<td>6.54</td>
</tr>
</tbody>
</table>

School of Information Systems
Conclusion

• Take into account of a **user’s current basket information** in making **personalized item recommendations**.

• Propose two models (**BFM & CBFM**) contribute **statistically significant** improvements over:
  – **Factorization Machine (FM)**
  – **Association Rule (ASR)**
in terms of top-K recommendations.
Thank you for listening!
Q&A
Backup slides
What is the relationship between Model Complexity & Response Time?

Half-life Utility and Response Time on BeiRen
Negative Tuple Sampling

For each positive tuple \( t = \langle u_i, B_i, v_j, 1 \rangle \), we sample two negative tuples \( t^- = \langle u_i, B_i^-, v_j^-, -1 \rangle \):

- As \( v_j^- \), we pick an item never selected by the user
- \( B_i^- \) contains items that never co-occur with either user, \( v_j^- \), and other current items in \( B_i^- \)