Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba

Jizhe Wang Pipei Huang Huan Zhao Zhibo Zhang Binqiang Zhao Dik Lun Lee

RS in Taobao

RS on Mobile Taobao App Homepage

Major challenges facing RS in Taobao

- **Scalability**: Existing recommender system work well on smaller scale datasets, they fail on the much larger scale dataset in Taobao which has 1 billion users and 2 billion items.

- **Sparsity**: Extremely difficult to train an accurate recommending model since the interactions between users and items are sparse.

- **Cold Start**: Millions of new items are continuously uploaded each hour in Taobao. It is challenging to process these items or predict the preferences of users for these items.

Proposed Framework

- **Graph Construction**: We construct item graph from users’ behaviors and filter out invalid data and abnormal behaviors to eliminate noise.

- **Sequence Generation**: By running random walk, we can generate a number of sequences.

- **Embedding Training**: Apply the Skip-Gram algorithm to learn the embeddings, which maximizes the co-occurrence probability of two commodities in the obtained sequences.

Visualization

- **System Deployment**
  - **ONLINE**
    - RSP
    - TPP
    - RE
  - **OFFLINE**
    - XTF
    - EM

Embedding with Side Information

- **Side-information**: In e-commerce, side information refers to the category, shop, purchase-level, material, etc.

- **GES**: Add a layer with average-pooling operation to incorporate side information.

\[ H_v = \frac{1}{n+1} \sum_{s=0}^{n} W_s v \]

- **EGES**: Different side information contribute differently to the co-occurrence of items in users’ behaviors.

\[ H_v = \sum_{j=0}^{n} e^{a_j} W_j \]

Experiments

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Amazon</th>
<th>Taobao</th>
</tr>
</thead>
<tbody>
<tr>
<td>BGE</td>
<td>0.9327</td>
<td>0.897</td>
</tr>
<tr>
<td>LINE(1st)</td>
<td>0.9554</td>
<td>0.9100</td>
</tr>
<tr>
<td>LINE(2nd)</td>
<td>0.8664</td>
<td>0.9411</td>
</tr>
<tr>
<td>GES</td>
<td>0.97/7%</td>
<td>0.97/14%</td>
</tr>
<tr>
<td>EGES</td>
<td>0.97/10%</td>
<td>0.97/10%</td>
</tr>
</tbody>
</table>

EXPERIMENTS

Dataset: Amazon Taobao

BGE 0.9327 0.897
LINE(1st) 0.9554(+2.43%) 0.9100(+3.44%)
LINE(2nd) 0.8664(-7.65%) 0.9411(+6.98%)
GES 0.97/7%(-2.86%) 0.97/14(+10.1%)
EGES 0.97/10%(-4.00%) 0.97/10%(+10.3%)