A^3NCF: An Adaptive Aspect Attention Model for Rating Prediction

Zhiyong Cheng¹, Ying Ding², Xiangnan He¹, Lei Zhu³,
Xuemeng Song⁴, Mohan Kankanhalli⁴

1. National University of Singapore
2. Vipshop US Inc., USA
3. Shandong Normal University, China
4. Shandong University, China
MOTIVATION

- **Review-based recommendation**: review contains rich information about user preference and item features.

### Traditional Models
- Weight on Quality of ratings, RecSys’12
- User reviews as content, RecSys’13
- Aspect Weighting, WI’15
- TriRank, CIKM’15

### Topic Models & Latent Factors
- HFT, RecSys’13
- RMR, RecSys’14
- CMR, CIKM’14
- TopicMF, AAAI’14
- JMARS, KDD’14
- FLAME, WSDM’15
- RBLT, IJCAI’16
- ITLFM, TKDE’16
- ALFM, WWW’18

### Joint Models of Aspects and Ratings
- JMARS, KDD’14
- EFM, SIGIR’14
- TRCF, IJCAI’13
- SULM, KDD’17

### Deep Learning Models
- ConvMF, RecSys’16
- DeepCONN, WSDM’17
- TransNet, RecSys’17
- D-Attn, RecSys’17
- NARRE, WWW’18

M. Chelliah & S. Sarkar, RecSys’17 tutorial
**MOTIVATION**

- **Limitation**: ignores the fact that “a user may place different importance to the various aspects of different items”

<table>
<thead>
<tr>
<th>Traditional Models</th>
<th>Topic Models &amp; Latent Factors</th>
<th>Joint Models of Aspects and Ratings</th>
<th>Deep Learning Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight on Quality of ratings, RecSys’12</td>
<td>HFT, RecSys’13</td>
<td>JMARS, KDD’14</td>
<td>ConvMF, RecSys’16</td>
</tr>
<tr>
<td>User reviews as content, RecSys’13</td>
<td>RMR, RecSys’14</td>
<td>EFM, SIGIR’14</td>
<td>DeepCONN, WSDM’17</td>
</tr>
<tr>
<td>Aspect Weighting, WI’15</td>
<td>CMR, CIKM’14</td>
<td>JMARS, KDD’14</td>
<td>TransNet, RecSys’17</td>
</tr>
<tr>
<td>TriRank, CIKM’15</td>
<td>TopicMF, AAAI’14</td>
<td>FLAME, WSDM’15</td>
<td>D-Attn, RecSys’17</td>
</tr>
<tr>
<td></td>
<td>JMARs, KDD’14</td>
<td>RBLT, IJCAI’16</td>
<td>NARRE, WWW’18</td>
</tr>
<tr>
<td></td>
<td>JMARs, KDD’14</td>
<td>ITLIFM, TKDE’16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALFM, WWW’18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
OUR MODEL - OVERVIEW
OUR MODEL – INPUT MODULE

- User/item identity: binary one-hot encoding
- Embedding layer -> identity representation

- User/item features: from the user/item’s review
- Topic model -> topic distribution as features
OUR MODEL – TOPIC MODEL

- \( K \): number of latent topics
- \( \theta_u \): user feature – topic distribution of user \( u \)
- \( \varphi_i \): item feature – topic distribution of item \( i \)
- \( \pi_u \): decide the current word \( w \) is drawn from \( \theta_u \) or \( \varphi_i \)
- \( w \): a word in the review
- \( z \): the latent topic of the word \( w \)

Assumption:
- ✓ A sentence in a review focuses on the same topic \( z \)
- ✓ When written a sentence, a user could comment from his own preferences \( \theta_u \) or from item’s characteristics \( \varphi_i \) : user-dependent parameter: \( \pi_u \)

Our model: mimics the processing of writing a review sentence
Goal: Estimate \( \theta_u \) and \( \varphi_i \)
OUR MODEL – FUSION MODULE

Input

Feature Fusion

Attention

Prediction

Text Review
User (u)

Item (i)

0 0 0 1

Embedding

User Embedding

Item Embedding

User Feature (θ_u)

Item Feature (φ_i)

Topic model

Topic model

Fusion

ReLU

Element-wise Product

Attention Network

MLP Layers

Regression

\( \hat{y}_{ui} \)
OUR MODEL – FUSION MODULE

- **Fusion**: embedded feature + review-based feature
  - *Concatenation, addition, element-wise product*

- **Relu fully-connected layer**: further increasing the interaction between the two types of features
OUR MODEL – ATTENTION MODULE

- $p_u$: $k$-dimensional user feature
- $q_i$: $k$-dimensional item feature
- **Rating prediction**: inner product of user-feature and item-feature
- **Attention weight vector** $a_{u,i,k}$: introduce an attention weight $a_{u,i,k}$ to a factor $k$ to indicate the importance of this factor of item $i$ with respect to user $u$
  - For a user $u$, the importance weight of the factors are different with respect to each item $i$

$$F = a_{u,i} \odot (p_u \odot q_i)$$

$F$: $k$-dimensional feature $\rightarrow$ rating prediction
OUR MODEL – ATTENTION MECHANISM

• How to estimate the attention weight
  • User preferences and item characteristics can be observed in reviews → $\theta_u$ and $\varphi_i$
  • $p_u$ and $q_i$ are the fusion feature for the final prediction
  • Concatenation of the four feature: $\theta_u, \varphi_i, p_u, q_i$

Attention mechanism:

$$\hat{a}_{u,i} = v^T ReLU(W_a[\theta_u; \varphi_i; p_u; q_i] + b_a)$$

$$a_{u,i,k} = \frac{\exp(\hat{a}_{u,i,k})}{\sum_{j=1}^{K} \exp(\hat{a}_{u,i,j})}$$
OUR MODEL – RATING PREDICTION

- The obtained feature is fed into fully connected layers (one layer in our experiments)
- Rating prediction: regression

\[ z_L = \sigma_L (W_L (\sigma_{L-1} (W_{L-1} \cdots \sigma_1 (W_1 F + b_1) + b_{L-1}) + b_{L-1}) + b_L) \]

\[ \hat{r}_{u,i} = W z_L + b \]
EXPERIMENTAL SETUP

- **Dataset:** Five sub-datasets in the Amazon product Review dataset and The Yelp Dataset 2017
- **Setting:** training:validation:testing = 8:1:1
- **Task:** Rating prediction
- **Metrics:** RMSE (the smaller the better)

<table>
<thead>
<tr>
<th>Datasets</th>
<th># users</th>
<th># items</th>
<th># ratings</th>
<th>Sparsity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baby</td>
<td>17,177</td>
<td>7,047</td>
<td>158,311</td>
<td>0.9987</td>
</tr>
<tr>
<td>Grocery</td>
<td>13,979</td>
<td>8,711</td>
<td>149,434</td>
<td>0.9988</td>
</tr>
<tr>
<td>Home &amp; Kitchen</td>
<td>58,901</td>
<td>28,231</td>
<td>544,239</td>
<td>0.9997</td>
</tr>
<tr>
<td>Garden</td>
<td>1,672</td>
<td>962</td>
<td>13,077</td>
<td>0.9919</td>
</tr>
<tr>
<td>Sports</td>
<td>31,176</td>
<td>18,355</td>
<td>293,306</td>
<td>0.9995</td>
</tr>
<tr>
<td>Yelp2017</td>
<td>169,257</td>
<td>63,300</td>
<td>1,659,678</td>
<td>0.9998</td>
</tr>
</tbody>
</table>
EXPERIMENTAL SETUP - COMPETITORS

- **BMF**: Matrix factorization (MF) with biased terms
- **HFT**: Use *a linking function* to connect the latent factors in MF (ratings) and LDA (reviews)
- **RMR**: Mixture of Gaussian (ratings) +LDA (reviews)
- **RBLT**: Use *a linear combination* of the latent factors in MF (ratings) and LDA (reviews)
- **TransNet**: Neural networks on user and item reviews for rating prediction
PERFORMANCE COMPRASIONS

• All better than BMF: indicating the importance of reviews in preference modeling
• Review-based methods
  – are relative more stable than BMF with the increase of \#factor;
  – can achieve relatively good performance with a small \#factor
• \textsuperscript{3}NCF is the best; > RBLT (2.9\%↑) and > TransNet (2.2\%↑), because it
  – applies more complicate interactions to integrate reviews and ratings via non-linear neural networks,
  – uses an attention mechanism to capture users’ attention weights on different aspects of an item.
EFFECTS OF ASPCT ATTENTION

- **Comparisons**
  - NCF: without *review-based feature* and *attention mechanism*
  - ANCF: with *review-based feature* but without *attention mechanism*

- **Results**
  - **ANCF > NCF**: (1) the effectiveness of using reviews in recommendation; and (2) our model on integrating review and rating information
  - **A³NCF > ANCF**: (1) user’s attentions are varied for different items; and (2) the effectiveness of our attention model
CONCLUSIONS

- Advocate the point that “a user may place different attentions to different items”
- Propose an attentive neural network to capture a user’s attention weight for different items
- Conduct experiments on benchmarking dataset to demonstrate our viewpoints and the effectiveness of the proposed model
Thanks!

Homepage: https://sites.google.com/view/zycheng
E-mail: zhiyong.cheng@nus.edu.sg or jason.zy.cheng@gmail.com