Co-Attentive Multi-Task Learning for Explainable Recommendation

Zhongxia Chen

4/29/2019
Background

• Personalized recommendation has become a major technique for helping users handle huge amounts of online content

• Recommender systems remain mostly black boxes

• Except for accuracy, there is a growing interest in model explainability

• Providing explanations can increase user trust, improve satisfaction, and persuade the users to buy or try an item
Motivation

• A fundamental question of explainable recommendation: How we balance accuracy and explainability for explainable recommendation?

• Most existing methods consider the two goals in separate steps or only focus on one of the goals
  • Post-hoc
  • Embedded
  • Simple jointly learning method
Related Works—Post-hoc

- Users $U$
- Items $V$

Recommendation model $f(u, v)$ → Recommended items $V'$ → Explanation Method → Explanation $z$

Explanations provided by different methods

- Pre-defined template
- Retrieved from explanations written by others
- Generated by RNNs

Example explanations:

(a) Post-hoc
- Item: Last Stand of the 300
- User interest: war, history, documentary

(a1) Alice and 7 of your friends like this.
(b) Embedded-F
(b1) Because you watched Spartacus, we recommend Last Stand of the 300.
(c) Embedded-S
(c1) You might be interested in documentary, on which this item performs well.
(d) Joint
(d1) I agree with several others that this is a good companion to the movie.
(e) Ours
(e1) This is a very good movie.

(c2) This is a very good documentary about the battle of thermopylae.
Related Works—Post-hoc

• Explain a black-box model after it is trained
  • Separately consider accuracy and explainability
  • Information embedded inside the recommendation models are ignored

• Pros and cons
  • Highly readable and persuasive
  • Not reflect model’s actual reasoning
  • Difficult to generate in non-social scenarios
  • Limited diversity
Related Works—Embedded

Users $U$

Items $V$

Explanation Method

Explanation $z$

Recommended items $V'$

Item: Last Stand of the 300

User interest: war, history, documentary

(a) Post-hoc
Because you watched Spartacus, we recommend Last Stand of the 300.

(b) Embedded-F
You might be interested in documentary, on which this item performs well.

(c) Embedded-S
I agree with several others that this is a good companion to the movie.

(d) Joint
This is a very good movie.

(e) Ours
This is a very good documentary about the battle of thermopylae.

Pre-defined template
Retrieved from explanations written by others
Generated by RNNs

Explanations provided by different methods
Related Works—Embedded

- Integrate the explanation process into the construction of the recommendation model
  - Retrieval-based
  - Consist of features or sentences
  - But only focus on recommendation accuracy
    - Explainability is not included in the optimization goal

- Issues
  - Difficult to guarantee the quality of the explanations
  - Fail to provide a highly personalized explanation when data is sparse
  - Legal issues (copyright)
Related Works—Joint

Item: Last Stand of the 300
User interest: war, history, documentary

(a) Post-hoc
Because you watched Spartacus, we recommend Last Stand of the 300.
I agree with several others that this is a good companion to the movie.

(b) Embedded-F
You might be interested in documentary, on which this item performs well.
This is a very good movie.

(c) Embedded-S
This is a very good documentary about the battle of thermopylae.

(d) Joint

(e) Ours

Pre-defined template
Retrieved from explanations written by others
Generated by RNNs

Explanations provided by different methods

\[(r - \hat{r})^2\]

\[\sum_{w \in S} \log p(w)\]
Related Works—Joint

- A simple jointly learning method
  - Only shares user/item latent representations

- Issues
  - The shared representations are not explainable and fail to provide explicit constraints on the explanations
  - User/item embeddings do not contain sufficient information about deep user-item interactions
  - Generated explanations are usually quite general
Contributions

• We propose a Co-Attentive Multi-task Learning (CAML) model that tightly couples the recommendation task and the explanation task
  • Design an encoder-selector-decoder architecture for multi-task learning based on cognitive psychology
  • Propose a hierarchical co-attentive selector to effectively control the cross knowledge transfer for both tasks by incorporating multi-pointer networks
  • Our method improves both explainability and accuracy
Problem Formulation

• Input
  • User set $U$, item set $V$
  • User reviews $D_{u,1}, ..., D_{u,l_d}$, item reviews $D_{v,1}, ..., D_{v,l_d}$
  • Concepts
    • a subset of words that correspond to important explicit features mentioned in the review

  This is a great little comedy with a catchy song.

• Output
  • Rating $r$
  • Linguistic explanation $Y = (y_1, y_2, ..., y_T)$
    • illustrates why user $u$ likes or does not like item $v$
Method

1. **Encoder (Encoding)**
   - User \( u \) Reviews \( D_{u,1}, D_{u,2}, D_{u,3}, D_{u,4} \)
   - Item \( v \) Reviews \( D_{v,1}, D_{v,2}, D_{v,3}, D_{v,4} \)

2. **Multi-Pointer Co-Attention Selector (Storage)**
   - User Implicit Factor \( h_u \)
   - Item Implicit Factor \( h_v \)
   - Review Embedding \( d_{u,1}, d_{u,2}, d_{u,3}, d_{u,4} \)
   - Item Embedding \( d_{v,1}, d_{v,2}, d_{v,3}, d_{v,4} \)

3. **Multi-Task Decoder (Retrieval)**
   - User Co-Attention Pointer Selector
   - Item Co-Attention Pointer Selector
   - Concepts \( X_u, X_v \)
   - Task 1: Rating Regression
     - \( r \)
   - Task 2: Explanation Generation
     - Excellent
     - sci
     - fi
     - \(<EOS>\)
     - GRU
     - GRU
     - GRU

4. **Concept-level Co-attention**
   - User Embedding \( e_u, d'_u, c'_u \)
   - Item Embedding \( c'_v, d'_v, e_v \)
   - \( \alpha \)
   - Gumbel Pooling
   - \( \beta \)
   - Gumbel Pooling

5. **Review-level Co-attention**
   - User Reviews \( D_{u,1}, D_{u,2}, D_{u,3}, D_{u,4} \)
   - Item Reviews \( D_{v,1}, D_{v,2}, D_{v,3}, D_{v,4} \)
   - \( a \)
   - \( b \)
   - \( \Phi \)

6. **Task 1: Rating Regression**
   - \( r \)
   - \( y_1, y_2, y_3, y_4 \)
   - Excellent
   - sci
   - fi
   - \(<EOS>\)
   - FM

7. **Task 2: Explanation Generation**
   - Multi-Task Decoder
     - Multi-Pointer Aggregation
     - \( h_u, e_u^{(1)}, e_u^{(2)}, e_u^{(3)} \)
     - \( e_v^{(1)}, e_v^{(2)}, e_v^{(3)} \)
     - Concepts
     - \( X_u \)
     - \( X_v \)
Encoder

- Word Encoder

- Review Encoder
  - \( d_{u,i} = \sum_{w \in D_{u,i}} w \)

- User/Item implicit factor Encoder
  - Complement explicit factors of user/item
Multi-Pointer Co-Attention Selector

• Model the cross knowledge transferred for the two tasks

• Advantages of the multi-pointer co-attention networks (MPCN)
  • Faster convergence than REINFORCE
  • Model deep level user-item interactions

• Extend the MPCN to hierarchically select reviews and then concepts
  • Review-level co-attention pointer
  • Concept-level co-attention pointer
  • Multi-pointer aggregation
Review-level co-attention pointer

• Review-level Co-attention
  • \( \phi_{i,j} = F(d_{u,i})^T W_d F(d_{v,j}) \)

• Max-Pooling
  • \( a_i = \max_{j=1,\ldots,l_d} \phi_{i,j} \)
  • \( b_j = \max_{i=1,\ldots,l_d} \phi_{i,j} \)
Review-level co-attention pointer

• Gumbel-Softmax
  • $q_i = \frac{\exp(\frac{a_i + g_i}{\tau})}{\Sigma_{j=1}^{nm} \exp(\frac{a_j + g_j}{\tau})}$
  • $g_i$ is the Gumbel noise

• Forward pass: Hard attention
  $$ z_i = \begin{cases} 1, & i = \arg \max_j (q_j), \\ 0, & \text{otherwise} \end{cases} $$

• Backward pass: continuous gradients
Concept-level co-attention pointer

- Expand review to concept level

- Concept-level Co-attention
  - \( \psi_{i,j} = F(c_{u,i})^T W_c F(c_{v,j}) \)

- Mean-Pooling and Gumbel-Softmax
Multi-pointer aggregation

• Run selector multiple times

• Aggregate latent embeddings
  • Non-linear layer
    \[ \tilde{e}_u = \sigma(W_p[e_u^{(1)}, ..., e_u^{(n_p)}] + b_p) \]
    \[ \tilde{e}_v = \sigma(W_p[e_v^{(1)}, ..., e_v^{(n_p)}] + b_p) \]

• Collect selected concepts
Decoder

• Rating prediction
  • Factorization machine
    \[ \mathcal{L}_r = \frac{1}{2|\Omega|} \sum_{(u,v) \in \Omega} (r - r_r)^2 \]

• Explanation Generation
  • RNN decoder
  • Concept relevance loss
    \[ \mathcal{L}_c = \frac{1}{|\Omega|} \sum_{(u,v) \in \Omega} \sum_{t=1}^{T} (\max_k (\tau_k \log a_{t,k})) \]
  • Negative log-likelihood loss
    \[ \mathcal{L}_n = \frac{1}{|\Omega|} \sum_{(u,v) \in \Omega} \sum_{t=1}^{T} (\log a_{t,y_t}) \]

• Joint learning
  \[ \mathcal{L} = \mathcal{L}_r + \lambda_c \mathcal{L}_c + \lambda_n \mathcal{L}_n + \lambda_l \|\Theta\|_2^2 \]
Experiments

• Datasets
  • Amazon Electronics, Movies&TV
  • Yelp

• Baselines
  • Explainability
    • Retrieval-based: Lexrank, NARRE, RLRec
    • Generative: NRT
  • Rating prediction
    • CF: PMF, NMF, SVD++
    • Neural: MPCN, NARRE, RLRec, NRT

• Metrics
  • Explainability: Bleu and ROUGE, Human evaluation
  • Rating prediction: RMSE

<table>
<thead>
<tr>
<th></th>
<th>Electronics</th>
<th>Movies&amp;TV</th>
<th>Yelp-2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>192,403</td>
<td>123,960</td>
<td>677,379</td>
</tr>
<tr>
<td>Items</td>
<td>63,001</td>
<td>50,052</td>
<td>84,693</td>
</tr>
<tr>
<td>Reviews</td>
<td>1,688,117</td>
<td>1,697,533</td>
<td>2,530,843</td>
</tr>
<tr>
<td>Concepts</td>
<td>652</td>
<td>791</td>
<td>1,004</td>
</tr>
</tbody>
</table>
### Explainability Results

#### Overall Performance

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Criteria</th>
<th>Retrieval</th>
<th>Generative</th>
<th>Ours</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LexRank</td>
<td>NARRE</td>
<td>RLRec</td>
<td>CAML-G</td>
</tr>
<tr>
<td>Electronics</td>
<td>BLEU</td>
<td>1.44</td>
<td>1.45</td>
<td>1.45</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>ROUGE-1</td>
<td>14.22</td>
<td>15.19</td>
<td>11.12</td>
<td>17.39</td>
</tr>
<tr>
<td></td>
<td>ROUGE-2</td>
<td>3.60</td>
<td>3.29</td>
<td>1.60</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td>ROUGE-L</td>
<td>13.70</td>
<td>13.28</td>
<td>9.70</td>
<td>15.71</td>
</tr>
<tr>
<td></td>
<td>ROUGE-SU4</td>
<td>4.38</td>
<td>5.25</td>
<td>3.13</td>
<td>5.97</td>
</tr>
<tr>
<td>Movies&amp;TV</td>
<td>BLEU</td>
<td>1.78</td>
<td>1.75</td>
<td>1.73</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>ROUGE-1</td>
<td>15.68</td>
<td>15.31</td>
<td>11.61</td>
<td>18.09</td>
</tr>
<tr>
<td></td>
<td>ROUGE-2</td>
<td>2.45</td>
<td>3.62</td>
<td>3.84</td>
<td>4.30</td>
</tr>
<tr>
<td></td>
<td>ROUGE-L</td>
<td>12.46</td>
<td>12.99</td>
<td>10.06</td>
<td>16.00</td>
</tr>
<tr>
<td></td>
<td>ROUGE-SU4</td>
<td>5.24</td>
<td>5.79</td>
<td>4.98</td>
<td>6.29</td>
</tr>
<tr>
<td>Yelp</td>
<td>BLEU</td>
<td>0.97</td>
<td>1.13</td>
<td>1.13</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>ROUGE-1</td>
<td>11.06</td>
<td>10.55</td>
<td>9.28</td>
<td>13.31</td>
</tr>
<tr>
<td></td>
<td>ROUGE-2</td>
<td>2.42</td>
<td>2.66</td>
<td>1.93</td>
<td>3.05</td>
</tr>
<tr>
<td></td>
<td>ROUGE-L</td>
<td>10.15</td>
<td>9.30</td>
<td>8.18</td>
<td>12.13</td>
</tr>
<tr>
<td></td>
<td>ROUGE-SU4</td>
<td>3.58</td>
<td>3.85</td>
<td>3.10</td>
<td>4.60</td>
</tr>
</tbody>
</table>
Explainability Results

- Human Evaluation
  - 3 assessors, 100 test cases

<table>
<thead>
<tr>
<th></th>
<th>Electronics</th>
<th>Movies&amp;TV</th>
<th>Yelp-2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>192,403</td>
<td>123,960</td>
<td>677,379</td>
</tr>
<tr>
<td>Items</td>
<td>63,001</td>
<td>50,052</td>
<td>84,693</td>
</tr>
<tr>
<td>Reviews</td>
<td>1,688,117</td>
<td>1,697,533</td>
<td>2,530,843</td>
</tr>
<tr>
<td>Concepts</td>
<td>652</td>
<td>791</td>
<td>1,004</td>
</tr>
</tbody>
</table>

(a) Fluency

(b) Usefulness
## Case Study

| Case 1. User interest: **horror, night, fun** | 1. If you are a **fan** of the 80’s, you’ll love this.  
2. Not the best of the old **horror** movies, but it’s still a good one.  
3. Nice to see the old classic **horror** movies.  
| **NRT** | I’ll admit it.  
**CAML** | I am a huge fan of **horror** movies, and this is one of my favorite movies.  
**Truth** | Remember when **horror** movies were **fun**?  
| **GT** | I remember watching this film when I was a 8 years **kid**, I was so terrified, i didn’t want to go to the bath alone! |
| Case 2. User interest: **humor, scenery, main character** | 1. This movie is a total waste of **time**.  
2. Save your **money**.  
3. You know the movie is a **joke**.  
| **NRT** | I love this series.  
**CAML** | If you like british **humor**, you will love this series.  
**Truth** | This is very british **humor**.  
| **GT** | Embarassingly painful is what this crap. |
| Case 3. User interest: **story line, cartoon, worth** | 1. As a **fan** of the phantom of the opera, I was very excited to see this movie.  
2. If you are a **fan** of the phantom, you will love this movie.  
3. The **story** was a bit rushed to the end.  
**GT** | What a great **cast**. |
| **NRT** | I enjoyed this series as much as the first one.  
**CAML** | I enjoyed this movie, the **animation** was great and the **story line** was very good.  
 | Great **price** and the **animation** was cool.  
| **Truth** |  |
### Accuracy Results

<table>
<thead>
<tr>
<th>Datasets</th>
<th>CF</th>
<th>Neural Models</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PMF</td>
<td>NMF</td>
<td>SVD++</td>
</tr>
<tr>
<td>Electronics</td>
<td>2.065</td>
<td>1.170</td>
<td>1.105</td>
</tr>
<tr>
<td>Movies&amp;TV</td>
<td>1.250</td>
<td>1.089</td>
<td>1.013</td>
</tr>
<tr>
<td>Yelp</td>
<td>1.829</td>
<td>1.290</td>
<td>1.193</td>
</tr>
</tbody>
</table>
Conclusion

• We propose a co-attentive multi-task learning which fully exploits the correlations between the recommendation task and the explanation task

• We propose an encoder-selector-decoder architecture and a hierarchical co-attentive selector to effectively control the cross knowledge transfer for both tasks

• Experiments show that our approach outperforms state-of-the-art baselines on both the accuracy of rating prediction and the quality of generated explanations
Thank You!