Paradigm for Pre-training and Transfer Learning in Recommender Systems

Speaker: Fajie Yuan (PI of REPL Lab)
Time: 2023/09/23
CONTENTS

01  ID Overlapping-based Transfer
02  Modality-based Transfer
03  LLM-based Transfer
ID Overlapping-based Transfer
ID Overlapping-based Transfer

PeterRec (SIGIR2020)

Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. SIGIR2020
ID Overlapping-based Transfer

PeterRec (SIGIR2020)

A: branch of plum
B: Tree of peach
C: insertion
D: grow together

Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. SIGIR2020
ID Overlapping-based Transfer

PeterRec (SIGIR2020)

How we do these insertion?

Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. SIGIR2020
PeterRec (SIGIR2020)

- The first work achieving transferable recommendation between domains
- Shared information is need for ID mapping between domains
A person has different roles to play in different life scenes! But all these roles may have some commonalities, such as personalization, habits, preference.
ID Overlapping-based Transfer

Conure (SIGIR2021)

Catastrophic Forgetting!

How Conure does:

Parameter Changes

Last hidden Vector Changes

One Person, One Model, One World: Learning Continual User Representation without Forgetting. SIGIR2021
ID Overlapping-based Transfer

Conure (SIGIR2021)

Model Comparison:

(a) Standard Transfer Learning
(b) PeterRec
(c) Conure
(d) multi-task learning (MTL)

One Person, One Model, One World: Learning Continual User Representation without Forgetting. SIGIR2021
Conure (SIGIR2021)

- The first work proposing lifelong learning in recommendation
- Shared information is still need
CLUE (ICDM2021)

Learning transferable user representations with sequential behaviors via contrastive pre-training. ICDM2021
Modality-based Transfer
Modality-based Transfer

**TransRec**

*The first Recommender System regime enabling effective transfer across modalities & domains!*

(a) Transferring Process of TransRec

(b) Transfer of Single Modality Methods

TransRec: Learning Transferable Recommendation from Mixture-of-Modality Feedback. 2022/06
Modality-based Transfer

TransRec

(a) Pre-training on Source Domain

(b) Fine-tuning on Target Domains

TransRec: Learning Transferable Recommendation from Mixture-of-Modality Feedback. 2022/06
**TransRec**

**Result:**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Modality</th>
<th>Metric</th>
<th>IDRNN</th>
<th>IDCNN</th>
<th>IDRec</th>
<th>TFS</th>
<th>TransRec</th>
<th>Improv.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN-mixed</td>
<td>Mixed</td>
<td>HR@5</td>
<td>0.0109</td>
<td>0.0112</td>
<td>0.0117</td>
<td>0.0249</td>
<td>0.0285</td>
<td>14.46%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@5</td>
<td>0.0063</td>
<td>0.0067</td>
<td>0.0068</td>
<td>0.0160</td>
<td>0.0177</td>
<td>10.63%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HR@10</td>
<td>0.0129</td>
<td>0.0195</td>
<td>0.0210</td>
<td>0.0428</td>
<td>0.0478</td>
<td>11.68%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>0.0062</td>
<td>0.0094</td>
<td>0.0100</td>
<td>0.0213</td>
<td>0.0239</td>
<td>12.21%</td>
</tr>
<tr>
<td>TN-video</td>
<td>Image</td>
<td>HR@5</td>
<td>0.0134</td>
<td>0.0159</td>
<td>0.0153</td>
<td>0.0208</td>
<td>0.0271</td>
<td>30.29%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@5</td>
<td>0.0093</td>
<td>0.0098</td>
<td>0.0092</td>
<td>0.0131</td>
<td>0.0173</td>
<td>30.06%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HR@10</td>
<td>0.0201</td>
<td>0.0265</td>
<td>0.0267</td>
<td>0.0336</td>
<td>0.0424</td>
<td>26.19%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>0.0114</td>
<td>0.0133</td>
<td>0.0125</td>
<td>0.0173</td>
<td>0.0221</td>
<td>27.75%</td>
</tr>
<tr>
<td>TN-text</td>
<td>Text</td>
<td>HR@5</td>
<td>0.0105</td>
<td>0.0123</td>
<td>0.0105</td>
<td>0.0303</td>
<td>0.0358</td>
<td>18.15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@5</td>
<td>0.0063</td>
<td>0.0078</td>
<td>0.0062</td>
<td>0.0192</td>
<td>0.0227</td>
<td>18.23%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HR@10</td>
<td>0.0189</td>
<td>0.0220</td>
<td>0.0192</td>
<td>0.0500</td>
<td>0.0597</td>
<td>19.40%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>0.0089</td>
<td>0.0109</td>
<td>0.0090</td>
<td>0.0255</td>
<td>0.0303</td>
<td>18.82%</td>
</tr>
<tr>
<td>DouYin</td>
<td>Image</td>
<td>HR@5</td>
<td>0.0059</td>
<td>0.0057</td>
<td>0.0023</td>
<td>0.0115</td>
<td>0.0146</td>
<td>26.96%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@5</td>
<td>0.0037</td>
<td>0.0035</td>
<td>0.0014</td>
<td>0.0073</td>
<td>0.0090</td>
<td>23.29%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HR@10</td>
<td>0.0096</td>
<td>0.0100</td>
<td>0.0035</td>
<td>0.0205</td>
<td>0.0259</td>
<td>26.34%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>0.0049</td>
<td>0.0049</td>
<td>0.0018</td>
<td>0.0101</td>
<td>0.0126</td>
<td>24.75%</td>
</tr>
</tbody>
</table>

a. TransRec performs consistently better than its training-from-scratch version, i.e., TFS.
b. TransRec performs better than ID-based methods as well.
MoRec (SIGIR2023)

MoRec vs. IDRec

IDRec’s weaknesses:

a. Fails when users and items have few interactions, e.g., the cold-start setting.
b. Pre-trained IDRec is not transferable across platforms given that userIDs and itemIDs are in general not shareable in practice.
c. Pure IDRec cannot benefit from the technical advances in other communities NLP and CV.

MoRec’s potential:

a. MoRec is comparable to or even surpass IDRec in no-cold setting.
b. MoRec build connections for RS and other communities and inherit their latest advances.
c. Pre-trained MoRec can transfer across domains even without shared information.
Modality-based Transfer

MoRec (SIGIR2023)

Illustration of IDRec vs MoRec. The only difference is the item encoder.
- IDRec uses an item ID embedding matrix.
- MoRec uses the pre-trained modality encoder.

Illustration of DSSM and SASRec.
Modality-based Transfer

MoRec (SIGIR2023)

Diverse pure modal-based dataset

(a) Item cases on ImageNet1K.
(b) Item cases on HM.
(c) Item cases on Bili.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Metrics</th>
<th>DSSM</th>
<th>SASRec</th>
<th>Improv.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IDRec</td>
<td>BERT\textsubscript{base}</td>
<td>RoBERT\textsubscript{a_base}</td>
</tr>
<tr>
<td>MIND</td>
<td>HR@10</td>
<td>3.58</td>
<td>2.68</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>NDCG@10</td>
<td>1.69</td>
<td>1.21</td>
<td>1.35</td>
</tr>
<tr>
<td>HM</td>
<td>HR@10</td>
<td>4.93</td>
<td>1.49</td>
<td>1.87</td>
</tr>
<tr>
<td></td>
<td>NDCG@10</td>
<td>2.93</td>
<td>0.75</td>
<td>0.94</td>
</tr>
<tr>
<td>Bili</td>
<td>HR@10</td>
<td>1.14</td>
<td>0.38</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>NDCG@10</td>
<td>0.56</td>
<td>0.18</td>
<td>0.27</td>
</tr>
</tbody>
</table>

MoRec vs IDRec (Regular Setting)

Accuracy with different pre-trained ME in MoRec.

Where to Go Next for Recommender Systems? ID- vs. Modality-based Recommender Models Revisited
Modality-based Transfer

NineRec

Existing datasets pose a major obstacle for Transferable Recommender Systems!

**E-commerce**
Modality-only TransRec is hard to learn on E-commerce dataset where price matters a lot

**Single Scenario**
Interaction from a single scenario suffer from semantic insufficiency for Transfer learning

**Pre-extracted Modality**
Large representation gap between RS and CV&NLP is difficult to remove

**Single Modality**
Single textual or visual information struggles to reflect the user preference

---

Title: Mike Tomlin: "Steelers 'accept responsibility' for role in brawl with Browns"
Category: Sport
NineRec

A large-scale benchmark dataset for exploring MoRec’s transferability between non-overlapping domains

- **Dataset scale:**
  - 1 source: #User: 2M, #Item: 140k
  - 9 targets: #User: 2k-20k, #Item: 1k-8k
- **No user overlap across targets**
- **Raw text and image**
- **Item from video platform:**
  - Interact mainly depend on content itself
  - Multiple targets across domain & platform
  - Diverse & semantically rich topics

Modality-based Transfer

PixelRec

A large-scale benchmark dataset for pure image-centric MoRec

- **Dataset scale:**
  - 200 million user-image interactions
  - 30 million users
  - 400,000 high-quality cover images

- **High-resolution raw image**
- **Rich features**
- **Diverse content topics**
Modality-based Transfer

Adapter-based TransRec

How to transfer in an efficient manner?

Modal-based transfer for downstream domains may heavy cost!
Modality-based Transfer

Adapter-based TransRec

*Only fine-tune Adapter networks when do transfer*
## Modality-based Transfer

### Adapter-based TransRec

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Architecture</th>
<th>Metrics</th>
<th>FTA</th>
<th>AdaT</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SASRec+BERT (Text)</td>
<td>HR@10</td>
<td>32.83</td>
<td>32.52</td>
<td>-0.94%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>17.33</td>
<td>17.44</td>
<td>+0.63%</td>
</tr>
<tr>
<td>MIND-&gt;Adressa</td>
<td>CPC+BERT (Text)</td>
<td>HR@10</td>
<td>29.56</td>
<td>30.07</td>
<td>+1.69%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>15.81</td>
<td>16.12</td>
<td>+1.92%</td>
</tr>
<tr>
<td>Trainable Parameters</td>
<td>100%</td>
<td>2.23%</td>
<td>-97.77%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SASRec+RoBERTa (Text)</td>
<td>HR@10</td>
<td>32.02</td>
<td>33.11</td>
<td>+3.38%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>16.95</td>
<td>17.54</td>
<td>+3.36%</td>
</tr>
<tr>
<td>H&amp;M-&gt;Amazon</td>
<td>CPC+RoBERTa (Text)</td>
<td>HR@10</td>
<td>29.90</td>
<td>30.64</td>
<td>+2.42%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>15.86</td>
<td>16.20</td>
<td>+2.16%</td>
</tr>
<tr>
<td>Trainable Parameters</td>
<td>100%</td>
<td>1.95%</td>
<td>-98.05%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SASRec+ViT (Image)</td>
<td>HR@10</td>
<td>29.00</td>
<td>27.66</td>
<td>-4.59%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>25.61</td>
<td>24.36</td>
<td>-4.88%</td>
</tr>
<tr>
<td></td>
<td>CPC+ViT (Image)</td>
<td>HR@10</td>
<td>26.56</td>
<td>25.29</td>
<td>-4.78%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>22.09</td>
<td>21.49</td>
<td>-2.72%</td>
</tr>
<tr>
<td>Trainable Parameters</td>
<td>100%</td>
<td>2.82%</td>
<td>-97.18%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SASRec+MAE (Image)</td>
<td>HR@10</td>
<td>28.10</td>
<td>25.67</td>
<td>-6.61%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>22.92</td>
<td>21.99</td>
<td>-4.05%</td>
</tr>
<tr>
<td></td>
<td>CPC+MAE (Image)</td>
<td>HR@10</td>
<td>27.50</td>
<td>25.18</td>
<td>-8.44%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDCG@10</td>
<td>23.31</td>
<td>21.83</td>
<td>-7.14%</td>
</tr>
<tr>
<td>Trainable Parameters</td>
<td>100%</td>
<td>2.82%</td>
<td>-97.18%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Text Scenario:** Comparable results but only 3% parameters fine-tuned

**Image Scenario:** Still worse than fine-tuning all parameters

---

LLM-based Transfer
The first task-agnostic pre-training framework in Recommender System

Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5). Recsys2022
**LLM-based Transfer**

**P5 (RecSys 2022)**

**ID-based Prompt Engineering:**

**Rating / Review / Explanation raw data for beauty**

- **user_id:** 7641
- **user_name:** stephanie
- **item_id:** 2001
- **item_title:** SHANY Nail Art Kit (24 Famous Colors Nail Art Polish, Nail Art Decoration)
- **review:** Absolutely great product. I bought this for my fourteen year old niece for Christmas and of course I had to try it out, then I tried another one, and another one and another one. So much fun! I even contemplated keeping a few for myself!
- **star_rating:** 5
- **summary:** Perfect!

**Sequential Recommendation raw data for beauty**

- **user_id:** 7641
- **user_name:** Vitorz
- **purchase_history:** 652 -> 466 -> 467 -> 653 -> 654 -> 655 -> 656 -> 657 -> 660 -> 668
- **next_item:** 552

**Direct Recommendation raw data for beauty**

- **user_id:** 256
- **user_name:** mariah nine
- **target_item:** 520
- **random_negative_item:** 9711

**Questions and Prompts**

- Which star rating will user_{{user_id}} give item_{{item_id}}? (1 being lowest and 5 being highest)
- Based on the feature word {{feature_word}}, generate an explanation for user_{{user_id}} about this product: {{item_title}}
- Give a short sentence describing the following product review from {{user_name}}: {{review}}
- Here is the purchase history of user_{{user_id}}: {{purchase_history}}. What to recommend next for the user?
- Choose the best item from the candidates to recommend for {{user_name}}? Use {{candidate_items}}.
Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5)
LLM-based Transfer

LLM4Rec

Does LLM enable emergent ability for Recommender System?

Exploring the Upper Limits of Text-Based Collaborative Filtering Using Large Language Models: Discoveries and Insights
LLM-based Transfer

LLM4Rec

**Q(i):** Does RS performance respond to the continuous increase in the item encoder’s size?

**Q(ii):** Can the 175B parameter LLM achieve universal text representation?

**Q(iii):** Can the 175B parameter LLM easily beat the simplest ID embedding based models (IDCF)?

**Q(iv):** How close is the LLMs to a universal recommendation model?

**Q(v):** Will recent prompt engineering based RS utilizing ChatGPT challenge MoRec with LLMs?

Exploring the Upper Limits of Text-Based Collaborative Filtering Using Large Language Models: Discoveries and Insights
LLM-based Transfer

Find our GitHub: westlake-repl

Updated collection about:

Pre-training and transfer learning in Recommender Systems

招实习！科研助理！博后！支撑大模型训练！
THANKS

Fajie Yuan
2023/09/23