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

Zheng Yuan¹, Fajie Yuan¹, Yu Song¹, Youhua Li¹, Junchen Fu¹, Fei Yang², Yunzhu Pan¹, Yongxin Ni¹

¹Westlake University; ²Zhejiang Lab
Background and Motivation

Framework and Training Details

Experiments and Findings

Future Works
Key weaknesses of ID–based Recsys (IDRec)
Background and Motivation

Key weaknesses of ID-based Recsys (IDRec)

(1) Cold-start setting
Background and Motivation

Key weaknesses of ID-based Recsys (IDRec)

(1) Cold-start setting

(2) Non-transferable
Background and Motivation

Key weaknesses of ID-based Reccsys (IDRec)

1. Cold-start setting
2. Non-transferable
3. Non-benefits

(1) Cold-start setting
(2) Non-transferable
(3) Non-benefits
Background and Motivation

Modality-based Recsys (MoRec)
Background and Motivation

Modality-based Recsys (MoRec)

(1) In past

ID embedding

Cold-start

Word2Vec, Glove ...
Background and Motivation

Modality-based Recsys (MoRec)

(1) In past

Word2Vec, Glove ...

ID embedding

Cold-start

(2) Now

BERT, GPT ...
ResNet, Swin ...

ID embedding

Regular

Cold-start
Background and Motivation

Modality-based Recsys (MoRec)

ID embedding

ID embedding

Word2Vec, Glove ...

ID embedding

ID embedding

ID1

ID2

Cold-start

Cold-start

Regular

Cold-start

Transferability

Benefits

Regular

(1) In past

(2) Now

BERT, GPT ...

ResNet, Swin ...
Background and Motivation

Rethink the potential of MoRec

Regular

Cold-start

Transferability

Benefits

Regular

ID embedding

BERT, GPT ...
ResNet, Swin ...
Background and Motivation

Rethink the potential of MoRec (1/4)
Rethink the potential of MoRec (1/4)

Q(i): Equipped with strong modality encoders, can MoRec be comparable to or even surpass IDRec in no-cold setting?
Rethink the potential of MoRec (2/4)
Q(ii): Can the recent technical advances developed in NLP and CV fields translate into improvement in MoRec?
Rethink the potential of MoRec (3/4)
Rethink the potential of MoRec (3/4)

Q(iii): How can we effectively use the item modality representations derived from a pre-training NLP or CV encoder?
Rethink the potential of MoRec (4/4)
Q(iv): Several key **challenges** that remain unexplored for MoRec training.
Overview

Background and Motivation

Framework and Training Details

Experiments and Findings

Future Works
Illustration of IDRec vs MoRec.
Datasets and Loss Function
Datasets and Loss Function

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

Item cases on datasets.
Datasets and Loss Function

Item cases on datasets.

Binary cross entropy loss

\[
\min - \sum_{u \in U} \sum_{i \in [2, \ldots, L]} \{ \log(\sigma(\hat{y}_{ui})) + \log(1 - \sigma(\hat{y}_{uj})) \} \quad \text{SASRec}
\]

\[
\min - \sum_{<u,i,j> \in R} \{ \log(\sigma(\hat{y}_{ui})) + \log(1 - \sigma(\hat{y}_{uj})) \} \quad \text{DSSM}
\]
Overview

Background and Motivation

Framework and Training Details

Experiments and Findings

Future Works
Q(i):

- Q(i): Equipped with strong modality encoders, can MoRec be comparable to or even surpass IDRec in no-cold setting?
Q(i): Equipped with strong modality encoders, can MoRec be comparable to or even surpass IDRec in no-cold setting?

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Metrics</th>
<th>IDRec</th>
<th>BERT&lt;sub&gt;base&lt;/sub&gt;</th>
<th>RoBERT&lt;sub&gt;base&lt;/sub&gt;</th>
<th>IDRec</th>
<th>BERT&lt;sub&gt;small&lt;/sub&gt;</th>
<th>BERT&lt;sub&gt;base&lt;/sub&gt;</th>
<th>RoBERT&lt;sub&gt;base&lt;/sub&gt;</th>
<th>Improv.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIND</td>
<td>HR@10</td>
<td>3.58</td>
<td>2.68</td>
<td>3.07</td>
<td>17.71</td>
<td>18.50</td>
<td>18.23</td>
<td>18.68</td>
<td>+5.48%</td>
</tr>
<tr>
<td></td>
<td>NDCG@10</td>
<td>1.69</td>
<td>1.21</td>
<td>1.35</td>
<td>9.52</td>
<td>9.94</td>
<td>9.73</td>
<td>10.02</td>
<td>+5.25%</td>
</tr>
<tr>
<td>HM</td>
<td>HR@10</td>
<td>4.93</td>
<td>1.49</td>
<td>1.87</td>
<td>6.84</td>
<td>6.67</td>
<td>6.97</td>
<td>7.24</td>
<td>+5.85%</td>
</tr>
<tr>
<td></td>
<td>NDCG@10</td>
<td>2.93</td>
<td>0.75</td>
<td>0.94</td>
<td>4.01</td>
<td>3.56</td>
<td>3.80</td>
<td>3.98</td>
<td>-0.75%</td>
</tr>
<tr>
<td>Bili</td>
<td>HR@10</td>
<td>1.14</td>
<td>0.38</td>
<td>0.57</td>
<td>3.03</td>
<td>2.93</td>
<td>3.18</td>
<td>3.28</td>
<td>+8.25%</td>
</tr>
<tr>
<td></td>
<td>NDCG@10</td>
<td>0.56</td>
<td>0.18</td>
<td>0.27</td>
<td>1.63</td>
<td>1.45</td>
<td>1.59</td>
<td>1.66</td>
<td>+1.84%</td>
</tr>
</tbody>
</table>

MoRec vs IDRec (Regular Setting)
Q(i): Equipped with strong modality encoders, can MoRec be comparable to or even surpass IDRec in no-cold setting?
Q(ii):

- Q(ii): Can the recent technical advances developed in NLP and CV fields translate into accuracy improvement in MoRec?
Q(ii): Can the recent technical advances developed in NLP and CV fields translate into accuracy improvement in MoRec?
Q(iii):

- Q(iii): How can we effectively use the item modality representations derived from an pre-training NLP or CV encoder?
Q(iii): How can we effectively use the item modality representations derived from an pre-training NLP or CV encoder?

<table>
<thead>
<tr>
<th>Dataset</th>
<th>IDRec</th>
<th>ME</th>
<th>TS</th>
<th>TS-DNN</th>
<th>E2E</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIND</td>
<td>17.71</td>
<td>BERT_{base}</td>
<td>13.93</td>
<td>16.66</td>
<td>16.32</td>
</tr>
<tr>
<td>HM</td>
<td>6.84</td>
<td>ResNet50</td>
<td>4.03</td>
<td>5.40</td>
<td>5.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Swin-T</td>
<td>3.45</td>
<td>4.46</td>
<td>5.28</td>
</tr>
<tr>
<td>Bili</td>
<td>3.03</td>
<td>ResNet50</td>
<td>0.72</td>
<td>1.23</td>
<td>1.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Swin-T</td>
<td>0.79</td>
<td>1.40</td>
<td>1.81</td>
</tr>
</tbody>
</table>

E2E vs TS with additional MLP layers
Q(iv):

• Q(iv): Several key challenges that remain unexplored for MoRec training.
Q(iv):

- Q(iv): Several key **challenges** that remain unexplored for MoRec training.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>#Param.</th>
<th>FLOPs</th>
<th>Time/E</th>
<th>MU</th>
<th>GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIND</td>
<td>IDRec</td>
<td>47M</td>
<td>0.12G</td>
<td>2.7m</td>
<td>3G</td>
<td>V100-32G(1)</td>
</tr>
<tr>
<td></td>
<td>BERT&lt;sub&gt;tiny&lt;/sub&gt;</td>
<td>11M</td>
<td>0.63G</td>
<td>10m</td>
<td>4G</td>
<td>V100-32G(1)</td>
</tr>
<tr>
<td></td>
<td>BERT&lt;sub&gt;small&lt;/sub&gt;</td>
<td>35M</td>
<td>16G</td>
<td>42m</td>
<td>13G</td>
<td>V100-32G(1)</td>
</tr>
<tr>
<td></td>
<td>BERT&lt;sub&gt;base&lt;/sub&gt;</td>
<td>116M</td>
<td>107G</td>
<td>102m</td>
<td>52G</td>
<td>V100-32G(2)</td>
</tr>
<tr>
<td>HM</td>
<td>IDRec</td>
<td>114M</td>
<td>1G</td>
<td>4.3m</td>
<td>5G</td>
<td>V100-32G(1)</td>
</tr>
<tr>
<td></td>
<td>ResNet18</td>
<td>18M</td>
<td>40G</td>
<td>95m</td>
<td>23G</td>
<td>V100-32G(1)</td>
</tr>
<tr>
<td></td>
<td>ResNet34</td>
<td>29M</td>
<td>81G</td>
<td>136m</td>
<td>30G</td>
<td>V100-32G(1)</td>
</tr>
<tr>
<td></td>
<td>ResNet50</td>
<td>31M</td>
<td>91G</td>
<td>83m</td>
<td>80G</td>
<td>V100-32G(4)</td>
</tr>
<tr>
<td></td>
<td>Swin-T</td>
<td>34M</td>
<td>96G</td>
<td>107m</td>
<td>157G</td>
<td>A100-40G(4)</td>
</tr>
<tr>
<td></td>
<td>Swin-B</td>
<td>94M</td>
<td>333G</td>
<td>102m</td>
<td>308G</td>
<td>A100-40G(8)</td>
</tr>
</tbody>
</table>

The training cost of IDRec and End2end MoRec
Q(iv):

- Several key **challenges** that remain unexplored for MoRec training.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>ME</th>
<th>TS</th>
<th>E2E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>w/o</td>
<td>w/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improv.</td>
<td></td>
</tr>
<tr>
<td>MIND</td>
<td>BERT_base</td>
<td>13.93</td>
<td>14.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+5.38%</td>
<td></td>
</tr>
<tr>
<td>HM</td>
<td>MAE_base</td>
<td>2.50</td>
<td>2.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+11.60%</td>
<td></td>
</tr>
<tr>
<td>Bili</td>
<td>MAE_base</td>
<td>0.57</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00%</td>
<td></td>
</tr>
</tbody>
</table>

Performing a second round of pre-training for ME.

(a) MoRec with BERT\_base on MIND.

(b) MoRec with Swin-T on HM.
- **Q(iv):** Several key **challenges** that remain unexplored for MoRec training.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>ME</th>
<th>IDRec</th>
<th>TS</th>
<th>ID+TS</th>
<th>Improv.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>w/o</td>
<td>w/</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ADD</td>
<td>CON</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ADD</td>
<td>CON</td>
<td></td>
</tr>
<tr>
<td>MIND</td>
<td>BERT&lt;sub&gt;base&lt;/sub&gt;</td>
<td>17.71</td>
<td>13.93</td>
<td>16.10</td>
<td>17.20</td>
<td>17.66</td>
</tr>
<tr>
<td>HM</td>
<td>Swin-T</td>
<td>6.84</td>
<td>3.45</td>
<td>5.75</td>
<td>4.89</td>
<td>5.37</td>
</tr>
<tr>
<td>Bili</td>
<td>Swin-T</td>
<td>3.03</td>
<td>0.79</td>
<td>3.01</td>
<td>2.61</td>
<td>3.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w/o</td>
<td>w/</td>
<td></td>
<td>w/o</td>
<td>Improv.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ADD</td>
<td>CON</td>
<td></td>
<td>ADD</td>
<td>CON</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ADD</td>
<td>CON</td>
<td></td>
<td>ADD</td>
<td>CON</td>
<td></td>
</tr>
<tr>
<td>16.66</td>
<td>14.93</td>
<td>16.58</td>
<td>17.29</td>
<td>17.55</td>
<td>-0.90%</td>
<td></td>
</tr>
<tr>
<td>5.55</td>
<td>5.27</td>
<td>4.00</td>
<td>4.77</td>
<td>5.11</td>
<td>-22.95%</td>
<td></td>
</tr>
<tr>
<td>2.10</td>
<td>2.86</td>
<td>2.35</td>
<td>2.50</td>
<td>2.72</td>
<td>-5.61%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18.23</td>
<td>16.25</td>
<td>17.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.97</td>
<td>5.40</td>
<td>4.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.18</td>
<td>2.94</td>
<td>2.55</td>
</tr>
</tbody>
</table>

**co-training ID and modality.**
Overview

Background and Motivation

Framework and Training Details

Experiments and Findings

Future Works
Future Works

1. Text, vision, voice and video
2. Multimodal MoRec
3. Improved networks;
4. Scale up training data
5. Large Language Models: ChatGPT, GPT4
Where to Go Next for Recommender Systems? 
ID- vs. Modality-based Recommender Models Revisited

Zheng Yuan¹, Fajie Yuan¹, Yu Song¹, Youhua Li¹, Junchen Fu¹,
Fei Yang², Yunzhu Pan¹, Yongxin Ni¹

Q&A