Directional and Explainable Serendipity Recommendation

Xueqi Li$^1$, Wenjun Jiang$^1$, Weiguang Chen$^1$, Jie Wu$^2$, Guojun Wang$^3$ and Kenli Li$^1$

$^1$Hunan University, $^2$Temple University and $^3$Guangzhou University

Presenter: Xueqi Li
Outline

◊ Introduction
◊ DESR: The details
◊ Experiments
◊ Conclusions & Future Work
Outline

◊ Introduction

◊ DESR: The details

◊ Experiments

◊ Conclusions & Future Work
An example

◊ Accuracy-oriented recommendation
◊ Serendipity-oriented recommendation
◊ Preference direction
◊ The necessity of directional recommendations
An example

◊ Accuracy-oriented recommendation
◊ Serendipity-oriented recommendation
◊ Preference direction
◊ The necessity of directional recommendations
Open Challenges

◊ What are the users’ preference directions?
◊ How to generate serendipitous recommendations with user preference direction?
◊ Why the items are recommended?
Outline

◊ Introduction

◊ DESR: The details

◊ Experiments

◊ Conclusions & Future Work
The Concepts

- **User Preference Direction**
  - from a short-term demand to a long-term preference

- **Serendipity Vector**
  - Direction: the corresponding user preference direction
  - Magnitude: the range for serendipity recommendations
Problem Definition

• Input:
  a user set $U$, an item set $I$, a rating matrix $M$ and a target user $u_t$

• Output:
  a list of potential items $I_t$ and the corresponding explanations

• Objective:
  Maximize serendipity($I_t, u_t$)
Framework

◊ Long-term Preference Extraction
◊ Short-term Demand Capture
◊ Recommendation Generation
◊ Back-routing for Explanations
Long-term Preference Extraction with GMM

$$P(y|\theta) = \sum_{k=1}^{K} \alpha_k \phi(y|\theta_k),$$

$$\phi(y|\theta_k) = \frac{1}{\sqrt{2\pi} \sigma_k} \exp\left(-\frac{(y - \mu_k)^2}{2\sigma_k^2}\right),$$

<table>
<thead>
<tr>
<th>Clustering</th>
<th>Preference Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_k$</td>
<td>k-th clustering center</td>
</tr>
<tr>
<td>$\sigma_k$</td>
<td>the corresponding clustering level</td>
</tr>
</tbody>
</table>
Short-term Demand Capture with Capsule Network

\[ \text{loss} = \frac{1}{n} \sum_{i=1}^{n} (\text{deCap}_i - \hat{\text{deCap}}_i)^2 \]
Determining the Magnitude

\[ S(u_{\text{tar}}) : \text{the scope of the long-term preferences} \]
\[ F(\text{preCap}) : \text{the familiarity for preCap} \]

\[ \| \text{serendipity vector} \| = m_{\text{base}} (1 + S(u_{\text{tar}}))(1 + F(\text{preCap})) \]
Recommendation Generation

Setting the direction

\[
\frac{\text{serendipity vector}}{\|\text{serendipity vector}\|} = \frac{\text{preCap} - \text{deCap}}{\|\text{preCap} - \text{deCap}\|}
\]

Recommendation Generation

\[
s_i = \frac{T_i}{\sum_{i=1}^{K} T_i}, \text{ } T_i \text{ is the number of items related to the corresponding preference}
\]
Back-routing for Explanations

\[ expCap = \text{Minimize } \text{dis}(\text{cap}, \text{reCap}_{\text{tar}}) \]

\[
\text{explanation} = \begin{cases} 
  \text{explanation}_1 & \text{if } \text{expCap} \in \text{preCaps} \\
  \text{explanation}_2 & \text{if } \text{expCap} \in \text{deCaps} 
\end{cases}
\]

"The item is similar to the items, \( \{t_1, t_2, ..., t_E\} \), which you watched for a long time."

"The item is similar to the items, \( \{t_1, t_2, ..., t_E\} \), which you recently watched."
Outline

◊ Introduction
◊ DESR: The details
◊ Experiments
◊ Conclusions & Future Work
Experiments

◊ Experiment Settings
◊ Effects of Parameters
◊ Overall Comparison
◊ Verification on Components
◊ A Case Study
## Experiment Settings -- Data Preprocessing

### Table 2: Statistics of datasets.

<table>
<thead>
<tr>
<th>Item</th>
<th>Statistic</th>
<th>MovieLens-1m</th>
<th>Amazon-Kindle-Store</th>
</tr>
</thead>
<tbody>
<tr>
<td># users</td>
<td></td>
<td>6040</td>
<td>3061</td>
</tr>
<tr>
<td># items</td>
<td></td>
<td>3260</td>
<td>6073</td>
</tr>
<tr>
<td># ratings</td>
<td></td>
<td>998539</td>
<td>132594</td>
</tr>
<tr>
<td>density $^5$</td>
<td></td>
<td>5.07%</td>
<td>0.71%</td>
</tr>
</tbody>
</table>
Experiment Settings -- Baselines

◊ RAND  a random-based method
◊ ACC_{LSTM}  an accuracy-oriented method
◊ KFN  a serendipity-oriented method
◊ HAES  a serendipity-oriented method
## Experiment Settings -- Metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>accuracy (Eq. 16)</td>
<td>-</td>
<td>a metric to measure the recommendation accuracy</td>
</tr>
<tr>
<td>accuracy_preference (Eq. 17)</td>
<td>acc_pre</td>
<td>a metric to measure the similarity of reCaps and preCaps</td>
</tr>
<tr>
<td>accuracy_demand (Eq. 18)</td>
<td>acc_de</td>
<td>a metric to measure the similarity of reCaps and deCaps</td>
</tr>
<tr>
<td>accuracy_all (Eq. 19)</td>
<td>acc</td>
<td>a metric to summarize accuracy, acc_pre and acc_de</td>
</tr>
<tr>
<td>diversity (Eq. 20)</td>
<td>div</td>
<td>a metric to measure the diversity of recommendations</td>
</tr>
<tr>
<td>difference (Eq. 21)</td>
<td>-</td>
<td>a metric to measure the difference between recommendations and utar’s history</td>
</tr>
<tr>
<td>difference_all (Eq. 22)</td>
<td>dif</td>
<td>a metric to summarize diversity and difference</td>
</tr>
<tr>
<td>AD (Eq. 23)</td>
<td>-</td>
<td>a metric to comprehensively measure acc and dif</td>
</tr>
</tbody>
</table>

\[
AD = \frac{acc \times dif}{acc + dif}
\]

An overall metric for recommendation serendipity.

More details refer to the paper.
Varying parameters exerts different impacts for most approaches, especially ACC\textsubscript{LSTM} and DESR.

A bigger weight of acc\_pre (i.e., (η, θ) = (1, 0.5)) makes DESR outperform ACC\textsubscript{LSTM} on acc.
Effects of Parameters -- in metrics

◊ Serendipity-oriented methods (i.e., KFN, HAES, DESR) perform better on dif when a bigger weight is assigned to difference.

◊ DESR performs poorer than HAES on dif.

Figure 7: Effects of parameters on dif. (para0: $\lambda = 0.5, \gamma = 1$; para1: $\lambda = 1, \gamma = 0.5$; para2: $\lambda = 1, \gamma = 1$.)
It is hardly possible to capture accurate short-term demands when L is too big or small.

The optimal value for L is 8.
By increasing of $m_{\text{base}}$, the performance on dif becomes better and that on acc becomes worse.

When $m_{\text{base}} = 0.4$, DESR reaches the best performance on AD.

$$\|\text{serendipity vector}\| = m_{\text{base}}(1 + S(utar))(1 + F(\text{preCap}))$$
Overall Comparison -- on Serendipity

◊ DESR achieves the best performance on both datasets.

◊ DESR performs better on MovieLens-1m.

<table>
<thead>
<tr>
<th>Method</th>
<th>acc@5</th>
<th>acc@10</th>
<th>dif@5</th>
<th>dif@10</th>
<th>AD@5</th>
<th>AD@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAND</td>
<td>0.19</td>
<td>0.19</td>
<td>0.777</td>
<td>0.874</td>
<td>0.152</td>
<td>0.156</td>
</tr>
<tr>
<td>ACC\textsubscript{LSTM}</td>
<td>0.477</td>
<td>0.469</td>
<td>0.348</td>
<td>0.543</td>
<td>0.201</td>
<td>0.252</td>
</tr>
<tr>
<td>KFN</td>
<td>0.298</td>
<td>0.28</td>
<td>0.37</td>
<td>0.669</td>
<td>0.165</td>
<td>0.198</td>
</tr>
<tr>
<td>HAES</td>
<td>0.308</td>
<td>0.304</td>
<td>0.595</td>
<td>0.732</td>
<td>0.203</td>
<td>0.215</td>
</tr>
<tr>
<td>DESR</td>
<td>0.491</td>
<td>0.48</td>
<td>0.466</td>
<td>0.66</td>
<td>0.239</td>
<td>0.278</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>acc@5</th>
<th>acc@10</th>
<th>dif@5</th>
<th>dif@10</th>
<th>AD@5</th>
<th>AD@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAND</td>
<td>0.237</td>
<td>0.236</td>
<td>0.715</td>
<td>0.805</td>
<td>0.178</td>
<td>0.183</td>
</tr>
<tr>
<td>ACC\textsubscript{LSTM}</td>
<td>0.566</td>
<td>0.558</td>
<td>0.338</td>
<td>0.455</td>
<td>0.212</td>
<td>0.25</td>
</tr>
<tr>
<td>KFN</td>
<td>0.269</td>
<td>0.295</td>
<td>0.355</td>
<td>0.582</td>
<td>0.153</td>
<td>0.196</td>
</tr>
<tr>
<td>HAES</td>
<td>0.282</td>
<td>0.281</td>
<td>0.644</td>
<td>0.739</td>
<td>0.196</td>
<td>0.203</td>
</tr>
<tr>
<td>DESR</td>
<td>0.514</td>
<td>0.51</td>
<td>0.412</td>
<td>0.562</td>
<td>0.229</td>
<td>0.267</td>
</tr>
</tbody>
</table>
The more accurate the recommendations are, the less diverse they become.

DESR maximizes diversity under the premise of ensuring recommendation accuracy.
Verification on Components -- Preference Extraction

Table 4: Comparison on original metrics. (ml-1m)

<table>
<thead>
<tr>
<th>method</th>
<th>accuracy@5</th>
<th>accuracy@10</th>
<th>acc_pre@5</th>
<th>acc_pre@10</th>
<th>acc_de@5</th>
<th>acc_de@10</th>
<th>div@5</th>
<th>div@10</th>
<th>difference@5</th>
<th>difference@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAND</td>
<td>0.15</td>
<td>0.15</td>
<td>0.128</td>
<td>0.129</td>
<td>0.291</td>
<td>0.291</td>
<td>0.739</td>
<td>0.831</td>
<td>0.819</td>
<td>0.918</td>
</tr>
<tr>
<td>ACC_{LSTM}</td>
<td>0.461</td>
<td>0.454</td>
<td>0.428</td>
<td>0.42</td>
<td>0.541</td>
<td>0.533</td>
<td>0.314</td>
<td>0.389</td>
<td>0.383</td>
<td>0.697</td>
</tr>
<tr>
<td>KFN</td>
<td>0.238</td>
<td>0.225</td>
<td>0.297</td>
<td>0.268</td>
<td>0.358</td>
<td>0.349</td>
<td>0.054</td>
<td>0.472</td>
<td>0.686</td>
<td>0.867</td>
</tr>
<tr>
<td>HAES</td>
<td>0.293</td>
<td>0.291</td>
<td>0.245</td>
<td>0.239</td>
<td>0.386</td>
<td>0.383</td>
<td>0.551</td>
<td>0.632</td>
<td>0.638</td>
<td>0.832</td>
</tr>
<tr>
<td>DESR</td>
<td>0.352</td>
<td>0.344</td>
<td><strong>0.683</strong></td>
<td><strong>0.668</strong></td>
<td>0.437</td>
<td>0.429</td>
<td>0.398</td>
<td>0.536</td>
<td>0.533</td>
<td>0.784</td>
</tr>
</tbody>
</table>

Table 5: Comparison on original metrics. (Kindle)

<table>
<thead>
<tr>
<th>method</th>
<th>accuracy@5</th>
<th>accuracy@10</th>
<th>acc_pre@5</th>
<th>acc_pre@10</th>
<th>acc_de@5</th>
<th>acc_de@10</th>
<th>div@5</th>
<th>div@10</th>
<th>difference@5</th>
<th>difference@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAND</td>
<td>0.222</td>
<td>0.221</td>
<td>0.177</td>
<td>0.177</td>
<td>0.311</td>
<td>0.31</td>
<td>0.683</td>
<td>0.768</td>
<td>0.748</td>
<td>0.843</td>
</tr>
<tr>
<td>ACC_{LSTM}</td>
<td>0.577</td>
<td>0.569</td>
<td>0.459</td>
<td>0.453</td>
<td><strong>0.661</strong></td>
<td><strong>0.652</strong></td>
<td>0.3</td>
<td>0.357</td>
<td>0.375</td>
<td>0.552</td>
</tr>
<tr>
<td>KFN</td>
<td>0.252</td>
<td>0.282</td>
<td>0.218</td>
<td>0.238</td>
<td>0.338</td>
<td>0.364</td>
<td>0.036</td>
<td>0.371</td>
<td>0.674</td>
<td>0.794</td>
</tr>
<tr>
<td>HAES</td>
<td>0.273</td>
<td>0.272</td>
<td>0.213</td>
<td>0.212</td>
<td>0.358</td>
<td>0.358</td>
<td>0.593</td>
<td>0.671</td>
<td>0.695</td>
<td>0.807</td>
</tr>
<tr>
<td>DESR</td>
<td>0.433</td>
<td>0.435</td>
<td><strong>0.624</strong></td>
<td><strong>0.609</strong></td>
<td>0.485</td>
<td>0.486</td>
<td>0.363</td>
<td>0.452</td>
<td>0.46</td>
<td>0.671</td>
</tr>
</tbody>
</table>

◊ DESR increases $acc_{pre}$ by at least 33.33% even when compared to ACC_{LSTM}.

◊ DESR performs even better than other serendipity-oriented methods on accuracy.
Capsule network and GRU reach the same optimal MSE on both datasets.

Capsule network has a faster convergence and outperforms GRU on explainability.
A Case Study – on Explanation Generation

For reCap0 (Almost Famous), "The movie is similar to the movies, Forrest Gump, Life Is Beautiful and American Beauty, which you watched for a long time."
Summary of Experiments

◊ GMM outperforms RNN on developing a comprehensive representation on long-term preferences.

◊ Capsule network has great potential in sequence processing.

◊ Keeping accuracy on users’ long-term preferences would improve the recommendation serendipity.

◊ In general, DESR achieves a better performance on AD.
Outline

◊ Introduction
◊ DESR: The details
◊ Experiments
◊ Conclusions & Future Work
Conclusions

◊ the reinforcement of user preference direction and explainability in serendipity recommendation

◊ Proposal for novel fine-grained metrics for serendipity

Future work

◊ provide more user-friendly explanations in the serendipity recommendation
Thanks!