User Novelty driven Personalized Item Recommendation

CHEN, Jun (陈俊)

PhD Graduate
Tsinghua University
Outline

- User Novelty driven Personalized Item Recommendation
  - User novelty classification
  - Repeat recommendation (with low user novelty)
  - Novel recommendation (with high user novelty)
  - Intransitive preference modeling (beyond a single angle of user novelty)
Background

- **In a conventional way to build RecSys**
  - Obtain user feedbacks, e.g., views/clicks/ratings.
  - Fit data with a model, e.g. CF, MF, LTR, NN.
  - Predict unseen items of potential user interests.
  - Constraints: time, social relations, geo locations, etc.

- **Task in this study**
  - Given user’s previous consumption history with timestamps, predict what (s)he will consume next.

  \[ \text{Predict } x_{ut} \quad \text{s.t. } \{ x_{u1}, x_{u2}, x_{u3}, ..., x_{u(t-1)} \} \]

- **Some potential methods**
  - Markov Chain model
  - Tensor factorization
  - Neural networks for sequence learning (GRU/LSTM)

- **Regardless of user’s intention on } x_{ut} \]
Background

- **In this study, user’s intention matters**
  - In psychology, the *openness* in the *Big Five Personality Traits* model describes people’s willingness to accept unknown things.
  - User’s selection is a reflection of their *intentions*: repeat familiar items or explore novel items.

- **Problem breakdown**
  A. User novelty classification
  B. Repeat recommendation (*low user novelty*)
  C. Novel recommendation (*high user novelty*)
  D. Intransitive preference modeling
A. User novelty classification

- **Motives**
  - Understand user’s intentions of consumptions.
  - Divide the candidate itemset for recommendations.

- **Problem**
  - Given the time window $W$ of user $u$’s consumption behaviors before time $t$, predict whether $u$ will choose the consumed items in $W$ at $t$.
  - User novelty: $P(x^u_t \notin W)$

- **Examples**
  - Given $u$’s recent playlist of songs $W$, predict if $u$ will repeat listening to a song in $W$ at next time step.
  - Given the sequence of $u$’s recent POI check-ins $W$, predict if $u$ will revisit a POI in $W$ at next time step.
A. User novelty classification

- In a binary classification scheme
- Feature extraction
  - A. Average normalized item popularity
  - B. Average normalized item reconsumption ratio
  - C. User reconsumption ratio
  - D. Window repeat ratio

\[ h_{IP}(v) = \frac{\ln(1 + freq(v))}{\max_{x \in V} \ln(1 + freq(x))} \quad h_{IP}(W_{u,t}) = \frac{1}{|W_{u,t}|} \sum_{v \in W_{u,t}} h_{IP}(v) \]

\[ h_{AIRR}(v) = \ln(1 + \frac{\sum_{u \in U} \sum_{x_i^u \in X^u} 1_{x_i^u = v \land x_i^u \in W_{u,t}}}{\sum_{u \in U} \sum_{x_i^u \in X^u} 1_{x_i^u = v}}) \quad h_{AIRR}(v) = \frac{h_{AIRR}(v)}{\max_{x \in V} h_{AIRR}(x)} \]

\[ h_{UIRR}(u) = \frac{\sum_{x_i^u \in X^u} 1_{x_i^u \in W_{u,t}}}{|X^u|} \]

\[ h_{WRR}(W_{u,t}) = 1 - \frac{|DS(W_{u,t})|}{|W_{u,t}|} \]
A. User novelty classification

- Feature significance

![Graphs showing feature significance](image)

User Novelty driven Personalized Item Recommendation
A. User novelty classification

- In a binary classification scheme
  - Classifiers

  Lasso classifier
  \[ \Pr_L(u, t) = w^T x_{u,t} \]
  \[ \arg\min_{w_L} L(w) = \sum_{u \in U} \sum_{t \in T_u} (w^T x_{u,t} - \mathbb{1}_{t \in W^u_k})^2 \]
  \[ \text{s.t.} \sum_i w_i = 1 \]

  Quadratic classifier
  \[ \Pr_Q(u, t) = \sqrt{w^T \text{diag}(x_{u,t})^2} w \]
  \[ \arg\min_{w_Q} Q(w) = \sum_{u \in U} \sum_{t \in T_u} (\sqrt{w^T \text{diag}(x_{u,t})^2} w - \mathbb{1}_{t \in W^u_k})^2 \]
  \[ \text{s.t.} w^T w = 1 \]

- Experiments
  - Publish the ManicTime dataset
  - Average 80% accuracy on Last.fm, Gowalla, BrightKite and ManicTime dataset.

B. Repeat recommendation

- **Motives**
  - Repeat consumptions are common
    - e.g. 77% in Last.fm [Kapoor, WSDM2015]
  - Some people have low user novelty at sometimes.
  - A lack of study in repeat recommendation.

- **Problem**
  - Precondition: $u$’s user novelty is low at time $t$.
  - Given the time window $W$ of user $u$’s consumption behaviors before time $t$, recommend items from $W$ to $u$ at time $t$.

- **Examples**
  - Recommend a preferred pop song listened several days before.
  - Recommend a nice steak house visited months before.

Indeed, it is a ranking problem with implicit user feedback!
B. Repeat recommendation

- **Bayesian Personalized Ranking**
  - *u’s preference on v, \( r_{uv} = \mathbf{u} \cdot \mathbf{v} \).*
  - Observed pairwise comparison \( v_i >_u v_j \).
  - Maximize: \( \Pi \sigma(\mathbf{u} \cdot \mathbf{v}_i - \mathbf{u} \cdot \mathbf{v}_j) \).
  - Estimate \( \mathbf{u} \) for all users, \( \mathbf{v} \) for all items.
  - SGD on log loss.

- **Time Sensitive**
  - User’s preference at each time step is different.
  - Reconsumed vs. non-consumed (temporal pairwise comparisons).
  - Tensor factorization may not work since the latent factors of the next step is unknown.

---

**How to capture user’s preference for repeat consumption?**
B. Repeat recommendation

- **Time-Sensitive Personalized Pairwise Ranking**
  - Incorporate time-sensitive features $f_{uvt}$:
    \[ r_{uvt} = u^\top v + u^\top A_u f_{uvt} = u^\top (v + A_u f_{uvt}) \]
  - $f_{uvt}$ is extracted from $u$’s previous interactions with $v$ in $W$.
  - $A_u$ is a personalized feature space transform matrix.
  - $r_{uvt}$ balances $u$’s static and temporal preference on $v$.
  - The probability of $u$ preferring $v_i$ to $v_j$:
    \[
    p(v_i >_{u t} v_j) = \sigma(r_{uv_it} - r_{uv_jt}) \\
    = \sigma(u^\top(v_i + A_u f_{uv_it} - v_j - A_u f_{uv_jt})) \\
    = \sigma(u^\top(v_i - v_j + A_u(f_{uv_it} - f_{uv_jt}))) \\
    = \frac{1}{1 + e^{-u^\top(v_i - v_j + A_u(f_{uv_it} - f_{uv_jt}))}}.
    \]
B. Repeat recommendation

- **Time-Sensitive Personalized Pairwise Ranking**
  - Objective (log loss with L-2 regularization):
    \[ J = \sum_{(u,v_i,v_j,t) \in D} -\ln p(v_i > v_j) + \frac{\lambda}{2} \sum_u \| A_u \|_F^2 + \frac{\nu}{2} (\| U \|_F^2 + \| V \|_F^2) \]
  - Training with SGD.
  - Negative sampling
    - For each repeat consumption on \( v_i \) at time \( t \), randomly sample \( s \) (e.g. 5) negative items \( v_j \) w.r.t. time window \( W \).
  - Recommend repeat items \( v \) by ranking with \( r_{uvt} \).
  - Codes: [https://github.com/chenjun082/ts-ppr](https://github.com/chenjun082/ts-ppr)

- **Time-sensitive Features**
  - Normalized item quality
  - Item reconsumption ratio
  - Recency feature
  - Dynamic familiarity
B. Repeat recommendation

- **Experiments**
  - Evaluation conducted on Last.fm and Gowalla datasets.

![Graphs showing MAAP for Top-1, Top-5, and Top-10 recommendations for Gowalla and Last.fm datasets.](image)

C. Novel recommendation

- **Motives**
  - Some people have *high user novelty* at sometimes.
  - One of the hottest topics in RecSys.
  - Building a memory-aware RecSys is less explored before.

- **Problem**
  - Given the time window $W$ of user $u$’s previous consumptions before time $t$, recommend items to $u$ at $t$ regardless of previous consumptions or not.

- **Main contributions**
  - Modeling user’s temporal interest on a given item as a process of *memory forgetting and enhancement* towards building an intelligent RecSys.
  - Simplify the variable-order Markov Chain model with a weighted first-order Markov Chain model.
C. Novel recommendation

- Memory forgetting as a human nature
  - Ebbinghaus forgetting curve
  - More forgetting curves

- Personalized memory forgetting patterns
  - Personalized forgetting speed
  - Personalized initial memory

Intelligent RecSys

User Novelty driven Personalized Item Recommendation
C. Novel recommendation

- **Interest-forgetting Markov model**
  - In a conventional variable-order Markov model, the probability of consuming $v$ at $t$ is the *conditional probability* of all previous consumptions.

  $$P(v|X_{u,t}^t) = P(v|x_{t-\Delta}^u, \ldots, x_{t-2}^u, x_{t-1}^u)$$

  - The proposed (simplified) model,

  $$P(v|X_{u,t}^t) = P(v|x_{t-\Delta}^u, \ldots, x_{t-2}^u, x_{t-1}^u) = \sum_{i=1}^{\Delta} P(v|x_{t-i}^u) \lambda(u, t, i)$$

  - The forgetting and enhancement of interest is implanted in $\lambda(u,t,i)$,

  $$\lambda(u, t, i) = \Phi(u, t, i) \Upsilon(u, t, x_{t-i}^u)$$
C. Novel recommendation

- **Interest forgetting**
  - To measure the interest retention $\Phi$.  
  - Log-linear, $\Phi = c_u \Delta t^{-\alpha_u}$
  - Exponential, $\Phi = c_u \Delta t^{-\alpha_u} e^{-\beta_u \Delta t}$
  - Hyperbolic, $\Phi = \frac{c_u}{\Delta t^{-\alpha_u}}$
  - Personalized parameters to estimate
    - $\alpha_u$, $\beta_u$, $c_u$

- **Interest enhancement**
  - To measure the accumulative interest
    - Logistic, $Y = \frac{2}{1 + e^{-\phi_u f(u,v,t)}}$
    - Rational, $Y = 1 + f(u, v, t) \phi_u$
  - Personalized parameters to estimate
    - $\phi_u$
C. Novel recommendation

- **One-step transition probability**
  - Maximum likelihood estimation
    - $P(v|x) = f(v, x) / f(x)$
  - Matrix factorization

$$P(v|X^{u, t}) = \sum_{i=1}^{\Delta} P(v|x_{t-i}^{u}) \lambda(u, t, i)$$

- **Objective (log loss)**

$$\Theta^* = \arg\min_{\Theta} \mathcal{L} = -\sum_{u} \sum_{\tilde{x}, X^{u, t}} \ln P(\tilde{x}|X^{u, t}; \Theta)$$

$$= -\sum_{u} \sum_{\tilde{x}, X^{u, t}} \ln \left( \sum_{i=1}^{\Delta} P(\tilde{x}|x_{t-i}^{u}) \Phi(u, t, i) \Upsilon(u, t, x_{t-i}^{u}) \right),$$

s.t. \( 1 \leq \Upsilon(u, t, x_{t-i}^{u}) \leq 2, \ 0 \leq \Phi(u, t, i) \leq 1, \ 0 \leq P(\tilde{x}|x_{t-i}^{u}) \leq 1. \)

- **GD training to get** \( \alpha_u, \beta_u, c_u \) and \( \phi_u. \)

- **Recommend by ranking with** \( P(v|X^{u, t}) \)
C. Novel recommendation

- **Experiments**
  - Evaluation conducted on Last.fm dataset.
  - Much higher accuracy in Top-$k$ recommendations.

D. Intransitive preference modeling

- **Motives**
  - The assumption of transitive preference is over-simplified sometimes.
  - Intransitive preference was observed in pairwise comparisons [Tversky, 1969].
  - There lacks computational models for intransitive preference.

- **Problem**
  - Given many pairwise comparisons \( \{v_i >_u v_j\} \) (intransitivity may exist), find a function \( f(u, v_x, v_y) \) that can predict if \( u \) prefers \( v_x \) over \( v_y \) or otherwise for an unseen triplet \( (u, v_x, v_y) \).

- **Impact of intransitive preference models**
  - Transitive preference models (RankNet [Burges, 2005], BPR [Rendle, 2009], MF [Koren, 2009]) fail due to the partial order of scalar preference values.
  - A more complex comparison schema is required.
  - Find the cause of intransitive preference.
D. Intransitive preference modeling

- **Major argument**
  - User has different judging criteria for different pairwise comparisons.
  - The different selections of judging criteria lead to intransitive preference.

- **Image as an example**
  - Judging criteria: colors, salience objects, theme, etc.
  - If comparing A and B using colors, comparing B and C using salience objects, and comparing A and C using image theme, the pairwise comparison results can be intransitive.

- What if more than one judging criterion are used in a single comparison?

<table>
<thead>
<tr>
<th>No.</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images</td>
<td><img src="image1.png" alt="Image A" /></td>
<td><img src="image2.png" alt="Image B" /></td>
<td><img src="image3.png" alt="Image C" /></td>
<td><img src="image4.png" alt="Image D" /></td>
</tr>
<tr>
<td>Color</td>
<td>Blue, Green, White</td>
<td>Green, Grey, White</td>
<td>Yellow, Orange, Red</td>
<td>Blue, Grey, White</td>
</tr>
<tr>
<td>Object</td>
<td>Mountain, Grass, Cow</td>
<td>Tree, Water, Rock</td>
<td>Sunset, Cloud, Sea</td>
<td>Sea, Tree, Beach</td>
</tr>
<tr>
<td>Theme</td>
<td>Country Life</td>
<td>Primitive Forest</td>
<td>Overseas Scenery</td>
<td>Beach View</td>
</tr>
</tbody>
</table>
D. Intransitive preference modeling

- Multi-Criterion (MuCri) Preference Models
  - Main idea: for a given (user, item) pair, compute multiple preference values with multiple criteria to form a preference vector.
  - Latent MuCri (L-MuCri) Model
    - Represent a/an user/item with multiple latent vectors (i.e. criteria).
      - $u = \{u_1, u_2, ..., u_D\}$, $h = \{h_1, h_2, ..., h_D\}$
    - The Top-$F$ criteria under which user $u$ prefers item $a$.
      $$\max_{1 \leq x \leq D_{lt}} u_x^\top h_x^a$$
    - The joint set of criteria that $u$ uses to compare item $a$ and $b$,
      $$C_{uab}^{lt} = (\max_{1 \leq x \leq D_{lt}} u_x^\top h_x^a) \cup (\max_{1 \leq y \leq D_{lt}} u_y^\top h_y^b)$$
    - $u$’s preference on item $a$ when comparing $a$ and $b$ (may change when $u$ compare $a$ and $c$ or others),
      $$R_{uab}^{lt} = \frac{1}{|C_{uab}^{lt}|} \sum_{x \in C_{uab}^{lt}} u_x^\top h_x^a.$$
D. Intransitive preference modeling

- **Multi-Criterion (MuCri) Preference Models**
  - The probability of user $u$ preferring $a$ to $b$ in pairwise comparison,

$$p_{lt}(a >_u b) = \frac{e^{r_{ua|b}^{lt}}}{e^{r_{ua|b}^{lt}} + e^{r_{ub|a}^{lt}}} = \frac{1}{1 + e^{-(r_{ua|b}^{lt} - r_{ub|a}^{lt})}}$$

- An illustration

[Diagram showing pairwise comparison and calculation of probabilities]
D. Intransitive preference modeling

- **Multi-Criterion (MuCri) Preference Models**
  - Objective,
    \[
    \mathcal{O} = \sum_u \sum_{(a,b) \in \mathcal{R}_u} -\ln p(a >_u b) + \frac{\lambda}{2} (\|\mathbf{U}\|_F^2 + \|\mathbf{H}\|_F^2) + \frac{\gamma}{2} \|\hat{\mathbf{U}}\|_F^2
    \]
  - Content-based MuCri (C-MuCri) Model
    - \(\mathbf{f}^a\) are the extracted content features of item \(a\).
    - For images, we extracted RGB, SIFT, GIST, LBP and deep features.
    \[
    C_{uab}^{ct} = (\arg\max_{1 \leq x \leq D_{ct}} \mathbf{u}_x^T \mathbf{f}^a_x) \cup (\arg\max_{1 \leq y \leq D_{ct}} \mathbf{u}_y^T \mathbf{f}^b_y)
    \]
  - Hybrid MuCri (H-MuCri) Model
    - Max: \(p(a >_u b) = \max(p_{lt}(a >_u b), p_{ct}(a >_u b))\)
    - Mean: \(p(a >_u b) = \frac{1}{2}(p_{lt}(a >_u b) + p_{ct}(a >_u b))\)
    - Product: \(p(a >_u b) = p_{lt}(a >_u b)p_{ct}(a >_u b)\)

- **Codes:** [https://github.com/chenjun082/mucri](https://github.com/chenjun082/mucri)
D. Intransitive preference modeling

- **Evaluation**
  - Publish a new pairwise scenery image comparison dataset (INRIA Holidays).
    - [https://github.com/chenjun082/holidays](https://github.com/chenjun082/holidays)
  - Accuracy performance

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Models</th>
<th>Accuracy 2-fold</th>
<th>Accuracy 5-fold</th>
<th>Accuracy 10-fold</th>
<th>AUC 2-fold</th>
<th>AUC 5-fold</th>
<th>AUC 10-fold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BRP [14]</td>
<td>0.7436±0.0047</td>
<td>0.7488±0.0088</td>
<td>0.7497±0.0152</td>
<td>0.7408±0.0044</td>
<td>0.7463±0.0092</td>
<td>0.7465±0.0158</td>
</tr>
<tr>
<td></td>
<td>GBPR [15]</td>
<td>0.7167±0.0134</td>
<td>0.7319±0.0094</td>
<td>0.7334±0.0192</td>
<td>0.7138±0.0123</td>
<td>0.7289±0.0109</td>
<td>0.7310±0.0196</td>
</tr>
<tr>
<td></td>
<td>BC-Inner [27]</td>
<td>0.7291±0.0030</td>
<td>0.7327±0.0119</td>
<td>0.7404±0.0130</td>
<td>0.7267±0.0030</td>
<td>0.7306±0.0127</td>
<td>0.7378±0.0118</td>
</tr>
<tr>
<td></td>
<td>BC-Dist [27]</td>
<td>0.7152±0.0050</td>
<td>0.7173±0.0038</td>
<td>0.7244±0.0193</td>
<td>0.7137±0.0058</td>
<td>0.7148±0.0035</td>
<td>0.7217±0.0194</td>
</tr>
<tr>
<td></td>
<td>L-MuCri</td>
<td>0.7622±0.0050</td>
<td>0.7829±0.0097</td>
<td>0.7832±0.0110</td>
<td>0.7602±0.0048</td>
<td>0.7809±0.0107</td>
<td>0.7813±0.0111</td>
</tr>
<tr>
<td></td>
<td>C-MuCri</td>
<td>0.7525±0.0056</td>
<td>0.7772±0.0072</td>
<td>0.7842±0.0107</td>
<td>0.7641±0.0064</td>
<td>0.7765±0.0172</td>
<td>0.7830±0.0099</td>
</tr>
<tr>
<td>INRIA Holidays</td>
<td>H-MuCri-Max</td>
<td>0.7535±0.0013</td>
<td>0.7686±0.0026</td>
<td>0.7728±0.0130</td>
<td>0.7520±0.0017</td>
<td>0.7672±0.0026</td>
<td>0.7690±0.0132</td>
</tr>
<tr>
<td></td>
<td>H-MuCri-Mean</td>
<td>0.7746±0.0071</td>
<td>0.7896±0.0010</td>
<td>0.7902±0.0101</td>
<td>0.7732±0.0071</td>
<td>0.7880±0.0096</td>
<td>0.7887±0.0113</td>
</tr>
<tr>
<td></td>
<td>H-MuCri-Prod</td>
<td>0.7690±0.0100</td>
<td>0.7927±0.0064</td>
<td>0.7905±0.0076</td>
<td>0.7758±0.0033</td>
<td>0.7858±0.0105</td>
<td>0.7890±0.0082</td>
</tr>
<tr>
<td></td>
<td>BRP [14]</td>
<td>0.7159±0.0018</td>
<td>0.7243±0.0072</td>
<td>0.7250±0.0112</td>
<td>0.7158±0.0018</td>
<td>0.7242±0.0072</td>
<td>0.7248±0.0112</td>
</tr>
<tr>
<td></td>
<td>GBPR [15]</td>
<td>0.7059±0.0004</td>
<td>0.7104±0.0034</td>
<td>0.7116±0.0030</td>
<td>0.7059±0.0003</td>
<td>0.7103±0.0034</td>
<td>0.7115±0.0030</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>BC-Inner [27]</td>
<td>0.6795±0.0029</td>
<td>0.6859±0.0024</td>
<td>0.6881±0.0051</td>
<td>0.6795±0.0028</td>
<td>0.6858±0.0025</td>
<td>0.6880±0.0050</td>
</tr>
<tr>
<td></td>
<td>BC-Dist [27]</td>
<td>0.6788±0.0009</td>
<td>0.6786±0.0018</td>
<td>0.6787±0.0067</td>
<td>0.6785±0.0010</td>
<td>0.6783±0.0018</td>
<td>0.6783±0.0067</td>
</tr>
<tr>
<td></td>
<td>L-MuCri</td>
<td>0.7789±0.0011</td>
<td>0.7997±0.0031</td>
<td>0.8052±0.0044</td>
<td>0.7788±0.0011</td>
<td>0.7997±0.0031</td>
<td>0.8052±0.0045</td>
</tr>
<tr>
<td></td>
<td>C-MuCri</td>
<td>0.7222±0.0002</td>
<td>0.7469±0.0079</td>
<td>0.7504±0.0081</td>
<td>0.7222±0.0002</td>
<td>0.7469±0.0079</td>
<td>0.7504±0.0081</td>
</tr>
<tr>
<td></td>
<td>H-MuCri-Max</td>
<td>0.6079±0.0006</td>
<td>0.6179±0.0032</td>
<td>0.6174±0.0111</td>
<td>0.6078±0.0006</td>
<td>0.6175±0.0030</td>
<td>0.6173±0.0010</td>
</tr>
<tr>
<td></td>
<td>H-MuCri-Mean</td>
<td>0.7748±0.0003</td>
<td>0.7937±0.0029</td>
<td>0.8005±0.0029</td>
<td>0.7747±0.0000</td>
<td>0.7937±0.0029</td>
<td>0.8005±0.0039</td>
</tr>
<tr>
<td></td>
<td>H-MuCri-Prod</td>
<td>0.7744±0.0002</td>
<td>0.7960±0.0021</td>
<td>0.8012±0.0048</td>
<td>0.7743±0.0002</td>
<td>0.7960±0.0020</td>
<td>0.8011±0.0048</td>
</tr>
</tbody>
</table>
D. Intransitive preference modeling

- **Evaluation**
  - Users’ top preferred latent or content criteria in pairwise image comparisons.
D. Intransitive preference modeling

- More information
  - [https://github.com/chenjun082/mucri](https://github.com/chenjun082/mucri)
  - [https://github.com/chenjun082/holidays](https://github.com/chenjun082/holidays)
User Novelty driven Personalized Item Recommendation

CHEN, Jun (陈俊)
PhD Graduate
Tsinghua University