From Ratings to Trust: An Empirical Study of Implicit Trust in Recommender Systems

Guibing Guo, Jie Zhang, Daniel Thalmann
Anirban Basu*, Neil Yorke-Smith**

School of Computer Engineering, NTU, Singapore
*KDDI R&D Laboratories, Inc., Japan
**American University of Beirut, Lebanon; and University of Cambridge, UK
Introduction

- Trust-based RSs
  - User–item ratings
  - User–user trust
  - Alleviating data sparsity, cold start, etc.
Introduction

- Trust types
  - Explicit
    - Binary trust only
    - Noisy trust: trusted users, different preferences
    - Sparse trust
  - Implicit
    - Inferred from user behaviors
    - Revealing implicit trust ties
    - Real values, richer information
Introduction

- **Trust Metrics**
  - Rating prediction of items only
  - No comparison with explicit trust

- **Our proposal**
  - Recover explicit relationships accurately
  - Reveal as much explicit trust as possible
Trust Definition

- Trust in RSs
  - one's belief towards the ability of others in providing valuable ratings

- Trust properties
  - Asymmetry
  - Transitivity
  - Dynamicity
  - Context dependency
Trust Metrics

- **TM1** (Lathia et al., 2008)
  
  \[ t_{u,v} = \frac{1}{|I_{u,v}|} \sum_{i \in I_{u,v}} \left( 1 - \frac{|r_{u,i} - r_{v,i}|}{r_{max}} \right) \]

- **TM2** (Yuan et al., 2010, Papagelis et al., 2005)
  
  \[ t_{u,v} = \begin{cases} 
  s_{u,v}, & \text{if } s_{u,v} > \theta, |I_{u,v}| > \theta_I \\
  0, & \text{otherwise} 
  \end{cases} \]
Trust Metrics

- **TM3** (Hwang and Chen, 2007)
  \[ p_{u,i} = r_u + (r_{v,i} - r_v) \]

- **TM3a**
  \[ t_{u,v} = \frac{1}{|I_{u,v}|} \sum_{i \in I_{u,v}} \left( 1 - \frac{|p_{u,i} - r_{u,i}|}{r_{\max}} \right) \]

- **TM3b**
  \[ t_{u,v} = \frac{|I_{u,v}|}{|I_u \cup I_v|} \left( 1 - \frac{1}{|I_{u,v}|} \sum_{i \in I_{u,v}} \left( 1 - \frac{|p_{u,i} - r_{u,i}|}{r_{\max}} \right)^2 \right) \]
Trust Metrics

- **TM4** (O'Donovan and Smyth, 2005)

Correct \( (r_{v,i}, r_{u,i}) \):  
\[ |p_{u,i} - r_{u,i}| < \epsilon \]

\[
t_{u,v} = \frac{|CorrectSet(v)|}{|RecSet(v)|}
\]
Trust Metrics

**TM5** (O'Donovan and Smyth, 2005)

\[
u_v = \frac{1}{|I_{u,v}|} \sum_{i \in I_{u,v}} \frac{|p_{u,i} - r_{u,i}|}{r_{\text{max}}}
\]

\[b_v = \frac{1}{2} (1 - u_v)(1 + s_{u,v})\]

\[d_v = \frac{1}{2} (1 - u_v)(1 - s_{u,v})\]

\[t_{u,v} = b_v\]
### Table 1: A comparison of different trust metrics in terms of trust properties

<table>
<thead>
<tr>
<th>Method</th>
<th>Asymm.</th>
<th>Transitive</th>
<th>Dynamic</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM1 [9]</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>TM2 [13, 18]</td>
<td>No</td>
<td>Yes, iff $s_{u,v} &gt; \theta_s$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>TM3a [7], TM3b [16]</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>TM4 [12]</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>TM5 [15]</td>
<td>No</td>
<td>Yes, iff $s_{u,v} &gt; \theta_s$</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
More about ratings
- Rating time
- Item category
- Rating noise
  - Assumption: ratings are accurate and reflecting users’ real preferences
Evaluation

- Experimental Settings
  - Two real-world datasets
  - 5-fold cross validation
  - Using suggested settings
    - TM1: $\theta_s = 0.07$, $\theta_I = 2$
    - TM3b: $\lambda = 0.15$
    - TM4: $\epsilon = 0.8$, or 1.5

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Users</th>
<th>Items</th>
<th>Ratings</th>
<th>Trust</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>FilmTrust</td>
<td>1,508</td>
<td>2,071</td>
<td>35,497</td>
<td>1,853</td>
<td>1.14%</td>
</tr>
<tr>
<td>Epinions</td>
<td>40,163</td>
<td>139,738</td>
<td>664,824</td>
<td>487,183</td>
<td>0.05%</td>
</tr>
</tbody>
</table>
Evaluation Metrics

- **Metrics for rating prediction**
  - $\text{MAE} = \frac{\sum_i |\hat{r}_i - r_i|}{N}$
  - $\text{RC} = \frac{P}{M}$

- **Metrics for trust ranking**
  - NDCG
  - Recall
Evaluation

- Performance of trust ranking

![Graphs showing performance metrics for FilmTrust and Epinions](image_url)
Evaluation

- Performance of rating prediction
### Performance of rating prediction

#### Table 3: Significance tests of the MAE differences relative to the CF method in FilmTrust

<table>
<thead>
<tr>
<th>Methods</th>
<th>Mean Diff.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM1 – CF</td>
<td>0.0055</td>
<td>9</td>
<td>9.075</td>
<td>7.978e-6</td>
</tr>
<tr>
<td>TM2 – CF</td>
<td>0.0258</td>
<td>9</td>
<td>12.926</td>
<td>4.077e-6</td>
</tr>
<tr>
<td>TM3a – CF</td>
<td>0.0162</td>
<td>9</td>
<td>14.254</td>
<td>1.756e-7</td>
</tr>
<tr>
<td>TM3b – CF</td>
<td>-0.0005</td>
<td>9</td>
<td>-0.357</td>
<td>0.729</td>
</tr>
<tr>
<td>TM4 – CF</td>
<td>0.0073</td>
<td>9</td>
<td>7.719</td>
<td>2.943e-5</td>
</tr>
<tr>
<td>TM5 – CF</td>
<td>0.0049</td>
<td>9</td>
<td>11.691</td>
<td>9.614e-7</td>
</tr>
</tbody>
</table>

#### Table 4: Significance tests of the MAE differences relative to the CF method in Epinions

<table>
<thead>
<tr>
<th>Methods</th>
<th>Mean Diff.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM1 – CF</td>
<td>-0.0324</td>
<td>9</td>
<td>-38.992</td>
<td>2.386e-11</td>
</tr>
<tr>
<td>TM2 – CF</td>
<td>0.0276</td>
<td>9</td>
<td>39.747</td>
<td>2.009e-11</td>
</tr>
<tr>
<td>TM3a – CF</td>
<td>-0.0284</td>
<td>9</td>
<td>-20.978</td>
<td>5.957e-9</td>
</tr>
<tr>
<td>TM3b – CF</td>
<td>0.0248</td>
<td>9</td>
<td>23.531</td>
<td>2.155e-9</td>
</tr>
<tr>
<td>TM4 – CF</td>
<td>-0.0320</td>
<td>9</td>
<td>-40.137</td>
<td>1.841e-11</td>
</tr>
<tr>
<td>TM5 – CF</td>
<td>-0.0031</td>
<td>9</td>
<td>-37.747</td>
<td>3.190e-11</td>
</tr>
</tbody>
</table>
Summary

- Two kinds of metrics show more performance measures
- Trust metrics relatively low
  - Explicit trust should be considered
  - Lack of consistency across datasets
  - Similarity-based metrics not work well
  - Similarity methods problematic themselves
Conclusion

- Studied 5 trust metrics
  - Properties of trust

- Proposed trust ranking metrics

- Conducted experiments
  - Trust metrics need improvement
Future Work

- Model-based approaches
- Utility comparison of explicit & implicit trust in predicting item ratings
- Enabling trust propagation
- More rating inform should be considered.
Thank You! & Questions?