A Probabilistic Model for Using Social Networks in Personalized Item Recommendation

Allison J.B. Chaney
Princeton University

Tina Eliassi-Rad
Rutgers University

David M. Blei
Columbia University

ajbc.io/spf
Personalized Item Recommendation

Anna Karenina
Winter’s Tale
East of Eden
???
Matrix Factorization

\[
\begin{align*}
\text{latent user preferences} & \approx \text{K} \\
\text{latent item attributes} & \approx \text{K}
\end{align*}
\]
Including Social Networks
Including Social Networks

• Matches our intuition
Including Social Networks

• Matches our intuition

• Introduces explainable serendipity
Including Social Networks

- Matches our intuition
- Introduces explainable serendipity
- Improves performance
Including Social Networks

- Matches our intuition
- Introduces explainable serendipity
- Improves performance
- Helps us learn about user behavior
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
An Example Etsy User
observed data

ratings

network

inference algorithm

model assumptions

learned parameters

item attributes

user preferences

user influence

recommendations

model assumptions

\[ \begin{align*}
\beta_i & \quad r_{ui} \\
\theta_u & \quad \tau_{uv} \\
\mu_p & \quad \mu_e \\
\mu_{\tau} & \end{align*} \]
Matrix Factorization
Matrix Factorization

\[ \beta_i \quad r_{ui} \quad \theta_u \]

\[ \mu_\beta \quad I \quad U \]

observed ratings
Matrix Factorization

Item attributes

User preferences

observed ratings
Social Poisson Factorization
Social Poisson Factorization

Item attributes

User preferences

User influence
\[ r_{ui} \mid r_{-u,i} \sim \text{Poisson} \left( \theta_u^\top \beta_i + \sum_{v \in N(u)} \tau_{uv} r_{vi} \right) \]
Posterior Inference:
How do we go from a generative model to finding the values of the variables that best fit our data?
The Posterior Distribution of the latent model parameters $\theta$ can be written as:

$$p(\beta, \theta, \tau | R, N, \mu) = \frac{p(\beta, \theta, \tau, R, N | \mu)}{\int_\beta \int_\theta \int_\tau p(\beta, \theta, \tau, R, N | \mu)}$$

This equation can be solved analytically when the observed data $R$ and model hyperparameters $\mu$ are easy to compute, but the integral on the right-hand side can be intractable.
Mean Field Variational Inference

intractable posterior
Mean Field Variational Inference

- easy to compute approximation
- intractable posterior
Recommendation

\[ E[r_{ui}] = E[\theta_u]^\top E[\beta_i] + \sum_{v \in N(u)} E[\tau_{uv}] r_{vi} \]
<table>
<thead>
<tr>
<th>source</th>
<th># users</th>
<th># items</th>
<th>% ratings</th>
<th>% edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ciao</td>
<td>7,000</td>
<td>98,000</td>
<td>0.038%</td>
<td>0.103%</td>
</tr>
<tr>
<td>Epinions</td>
<td>39,000</td>
<td>131,000</td>
<td>0.012%</td>
<td>0.011%</td>
</tr>
<tr>
<td>Flixster</td>
<td>132,000</td>
<td>42,000</td>
<td>0.122%</td>
<td>0.006%</td>
</tr>
<tr>
<td>Douban</td>
<td>129,000</td>
<td>57,000</td>
<td>0.221%</td>
<td>0.016%</td>
</tr>
<tr>
<td>Social Reader</td>
<td>122,000</td>
<td>6,000</td>
<td>0.065%</td>
<td>0.001%</td>
</tr>
<tr>
<td>Etsy</td>
<td>40,000</td>
<td>5,202,000</td>
<td>0.009%</td>
<td>0.300%</td>
</tr>
</tbody>
</table>

etsy.com and librec.net/datasets.html
Existing Methods for Including Social Networks

**SoRec**  

**RSTE**  
Ma et al., Learning to Recommend with Social Trust Ensemble, SIGIR 2009.

**SocialMF**  

**TrustMF**  
Yang et al., Social Collaborative Filtering by Trust, IJCAI 2013.

**TrustSVD**  
Guo et al., TrustSVD: Collaborative Filtering with Both the Explicit and Implicit Influence of User Trust and of Item Ratings, AAAI 2015.

librec.net
Evaluation on held-out data

\[ CRR(\text{user}) = \sum_{n=1}^{N} \frac{1[\text{rec}_n \in \mathcal{H}]}{n} = \sum_{i \in \mathcal{H}} \frac{1}{\text{rank}(i)} \]

\[ NCRR(\text{user}) = \frac{CRR(\text{user})}{\text{ideal CRR(\text{user})}} \]
Summary

• SPF performs better than comparison models
• SPF is interpretable and has explainable serendipity
• SPF scales well to large data
• Source code available at ajbc.io/spf
Thank you! Questions and suggestions welcome.

Thank you to Blei Lab colleagues and Guibing Guo (LibRec creator)

ajbc.io/spf