Introduction

- Applied Probabilistic Matrix Factorization (PMF) [1] and variations of PMF to the task of music recommendation;
- Developed Constrained Kernelized Probabilistic Matrix Factorization (cKPMF) which we show to be superior to all other models discussed in this project for our task at hand.
- Motivation: learn and experiment with a simple but effective framework for making recommender systems, that can deal with sparse and imbalanced rating data.
- Input: log(listening count), user network, artist tag assignment.
- Output: missing ratings (listening count)

Data

- hetrec2011-lastfm-2k [2], a set of social networking, tagging, and music artist listening information from last.fm.

User-artist listening count

<table>
<thead>
<tr>
<th>Coldplay</th>
<th>Molby</th>
<th>Gorillaz</th>
</tr>
</thead>
<tbody>
<tr>
<td>13883</td>
<td>8983</td>
<td>100</td>
</tr>
</tbody>
</table>

Artist Side information

<table>
<thead>
<tr>
<th>ballad</th>
<th>pissbass</th>
<th>chillout</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

User Side information

<table>
<thead>
<tr>
<th>users</th>
<th>Coldplay</th>
<th>Molby</th>
<th>Gorillaz</th>
</tr>
</thead>
<tbody>
<tr>
<td>543</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

log(listening count) as Pseudo-Ratings

Challenges

- sparsity
- Imbalance

Models

PMF

U, V both Sampled Row-wise

Assume generated from zero mean Normal Distribution

Constrained PMF [1]

Constrain U with W: users who rated the same artists should have similar latent feature vector

Kernelized PMF [3]

U, V both Sampled Column-wise.

Assume generated from zero-mean GP

Capture covariance between rows

Our model: cKPMF

U is sampled row-wise

V is sampled column-wise

MAP: Maximize log-posterior over \( Y, W, V \)

Experiments/Results

- 10% as validation; 10% as test; (5662 ratings)
- Training set 1: 80% of ratings (45296 ratings)
- Training set 2: 20% of ratings (11324 ratings)
- Training using Stochastic Gradient Descent
- Grid search for hyperparameters (latent dimension, learning rate, regulation coefficient) using validation set.
- Evaluation:

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i} (Y_{i} - \hat{Y}_{i})^2}
\]

% improvement in test RMSE over PMF by user group

Advantages

- Having more training data is more important than picking the best model. The best model performance with 20% ratings (1.224) is still worse than the worst model performance with 80% ratings (1.139).
- Kernelized models able to use side information to make predictions when ratings are sparse. Though both useful, user side information is not as effective as item side information.
- Difference in model performance is much more obvious when data is sparse. (with 20% ratings)
- All PMF variations made more improvement over the baseline PMF model for users with 0-5 ratings.
- Our cKPMF model exceed the performance of all others. Advantage is more obvious for infrequent users.

Future work

- Explore the effect of adding user and artist bias into all these models.
- Investigate the computational efficiency and convergence behaviour of these models in different settings.
- Explore Bayesian Probabilistic Matrix Factorization

References:

https://drive.google.com/open?id=1rt7CFxppZGFxF2pttmysw3FMSOm0dazJ