Latent Space Modeling in Collaborative Filtering

Peter Tiňo
School of Computer Science
University of Birmingham, UK

Joint work with:
Gabriela Polčicová
Faculty of Informatics and Information Technologies
Slovak University of Technology, Bratislava, Slovakia
Some motivations

- The amount of available information is steadily increasing ...
- Need automatic methods for making some sense of this ...
- When deciding e.g. which book to read/film to watch, people often rely on advise given by other people.
- This is possible only inside small communities, where people know other peoples’ interests.
- Automate the process of sharing evaluations and making recommendations among people that potentially do not know each other but have some ‘common tastes’.
- Collaborative filtering (CF) – recommendations are produced by leveraging the existing preferences.
Collaborative filtering

Two main approaches:

- **Memory-based:**
  To recommend a new item to a particular user, we take into account ratings from people with ‘similar’ interests.

- **Model-based:**
  Uses probabilistic modeling to infer new recommendations.

Much work has focused on designing accurate and fast algorithms for predicting ratings of items of interest.

Few systems for understanding and visualization of principal user preference patterns/rating trends buried inside large and sparse body of collaborative filtering data.
Latent class models for user ratings

[Hofmann; 2001], [Hofmann, Puzicha; 1999]

Three sets we will work with:

- the set $U$ of users,
- the set of films (items), $Y$,
- the set $V$ of rating values that are used by users to evaluate the items.

Predict the rating $v_{u,y} \in V$ given by a user $u \in U$ to a film $y \in Y$.

With each triplet $(u, y, v_{u,y})$ associate a latent variable (class) $z_{u,y} \in \mathcal{Z} = \{1, 2, \ldots, K\}$ that ‘explains’ why the user $u$ rates the film $y$ by $v_{u,y}$. 
The actual models

- **Type I:**
  Given a film $y$, all users from class $z$ tend to adopt the same (class-specific) rating pattern $P(v|y, z)$. Given a user $u$ and a film $y$, the probability of vote $v$ is modeled as

  \[
  P(v|y, u) = \sum_{z \in Z} P(v|y, z)P(z|u).
  \]

  $P(z|u)$ is the probability that the user $u$ ‘participates’ in class $z$.

- **Type II:**
  All users from class $z$ tend to adopt the same preferences over the [rating, film] pairs $(v, y)$. Given a user $u$,

  \[
  P(v, y|u) = \sum_{z \in Z} P(v, y|z)P(z|u).
  \]
Introducing a topology into latent classes

Information theoretic interpretation of SOM: Link between SOM and vector quantization through the notion of a noisy communication channel.

Topological organization of codebook vectors emerges through non-uniformity of the channel noise: to minimize the average quantization error at the receiving end of the communication channel, the codebook vectors that are more likely to be corrupted into each other should represent ‘similar’ data patterns, i.e. should lie ‘close’ to each other in the data space.
Grid topology of latent classes

Latent classes organized on a regular two-dimensional grid embedded in $[-1,1]^2 \subset \mathbb{R}^2$.

‘Similar’ classes will lie close to each other on the grid.

Chanel noise is expressed through the ‘neighborhood function’: probability of corrupting $z_1$ into $z_2$

$$P(z_2|z_1) = \frac{\exp\left(\frac{-\|z_1-z_2\|^2}{2\sigma^2}\right)}{\sum_{z \in \mathcal{Z}} \exp\left(\frac{-\|z_1-z\|^2}{2\sigma^2}\right)}$$

$\sigma > 0$ determines ‘specificity’ of the topological neighborhood of class $z_1$. 

P. Tiňo and G.Polčicová
User-conditional generative model

Two copies $\mathcal{Z}_Y$ and $\mathcal{Z}_Z$ of the latent space $\mathcal{Z}$.

For each user $u \in U$:

1. Generate a latent class index $z_Y \in \mathcal{Z}_Y$ by sampling $P(\cdot \mid u)$ on $\mathcal{Z}_Y$.

2. Instead of using $z_Y$ to index the ‘common interest patterns’ $P(v \mid y, z_Y)$ (type I) or $P(v, y \mid z_Y)$ (type II), transmit the class identification $z_Y$ through a noisy communication channel, and receive (a possibly different) class index $z_Z \in \mathcal{Z}_Z$ with probability $P(z_Z \mid z_Y)$.

3. Randomly generate a film-conditional rating $v$ with probability $P(v \mid y, z_Z)$ (type I) or a pair $(v, y)$ with probability $P(v, y \mid z_Z)$ (type II).
**Model formulation**

The models for user ratings have now the form

\[
P(z_Z|u) = \sum_{z_Y \in Z_Y} P(z_Z|z_Y)P(z_Y|u),
\]

\[
P(v|y,u) = \sum_{z_Z \in Z_Z} P(v|y,z_Z)P(z_Z|u) \quad \text{[type I]}
\]

\[
P(v,y|u) = \sum_{z_Z \in Z_Z} P(v,y|z_Z)P(z_Z|u) \quad \text{[type II]}
\]

Denote by \( \rho(v,y,z_Z) \) the probabilities \( P(v|y,u) \) and \( P(v,y|u) \) in type I and type II models, respectively.
Parameterisations of rating distributions $\rho(v, y, z_Z)$

- **Type I**
  - multinomial $P(v|y, z_Z)$ (I-Mult)
  - binomial $P(v|y, z_Z)$ (I-Bin)

- **Type II**
  - joint multinomial $P(v, y|z_Z)$ (II-Mult)
  - $P(v, y|z_Z) = P(v|z_Z)P(y|z_Z)$
    * both $P(y|z_Z)$ and $P(v|z_Z)$ are multinomials (II-IndM)
    * $P(y|z_Z)$ is multinomially and $P(v|z_Z)$ is binomially distributed (II-IndB)
Parameter estimation

Given a set $\mathcal{D} = \{(u_1, y_1, v_1), \ldots, (u_N, y_N, v_N)\}$ of $N$ observation triplets $(u, y, v_{u,y})$, the log likelihood of the model $\mathcal{D}$ is

$$
\mathcal{L} = \sum_u \sum_{y \in \mathcal{Y}_u} \log P(v_{u,y}|y,u) \quad \text{[type I]}
$$

$$
\mathcal{L} = \sum_u \sum_{y \in \mathcal{Y}_u} \log P(v_{u,y}, y|u) \quad \text{[type II]},
$$

where $\mathcal{Y}_u$ is the set of films evaluated by the user $u$.

Fit the model parameters $P(z_y|u)$ and $\rho(v, y, z_z)$ via the Expectation-Maximization (EM) algorithm.
EM algorithm - E step

Compute the expected values of latent variables (given a data item) using the current values of the model parameters:

\[
P(z_Y \mid y, u, v) = \frac{P(z_Y \mid u) \sum_{z_Z} \rho(v, y, z_Z) P(z_Z \mid z_Y)}{\sum_{z'_Y} P(z'_Y \mid u) \sum_{z_Z} \rho(v, y, z_Z) P(z_Z \mid z'_Y)},
\]

\[
P(z_Z \mid y, u, v) = \frac{\rho(v, y, z_Z) \sum_{z_Y} P(z_Z \mid z_Y) P(z_Y \mid u)}{\sum_{z'_Z} \rho(v, y, z'_Z) \sum_{z_Y} P(z'_Z \mid z_Y) P(z_Y \mid u)}.
\]
EM algorithm - M step

Re-estimate the model parameters by maximizing the expected complete data log-likelihood evaluated in the E-step.

Update of $P(z_Y|u)$ is the same for all types of models:

$$P(z_Y|u) = \frac{\sum_{y \in \mathcal{Y}_u} P(z_Y|y, u, v_{u, y})}{|\mathcal{Y}_u|}$$
M step - Multinomial $\rho(v, y, zz)$

$$
\rho(v, y, zz) = P(v|y, zz) = \frac{\sum_{u \in U_{v,y}} P(zz|y, u, v)}{\sum_{v'} \sum_{u \in U_{v',y}} P(zz|y, u, v')} \quad \text{[type I]}
$$

$$
\rho(v, y, zz) = P(v, y|zz) = \frac{\sum_{u \in U_{v,y}} P(zz|y, u, v)}{\sum_{v', y'} \sum_{u \in U_{v',y'}} P(zz|y', u, v')} \quad \text{[type II]}
$$

where $U_{v,y}$ is the set of users that evaluated film $y$ with rating $v$.

Factorised:

$$
P(y|zz) = \frac{\sum_{u \in U_y} P(zz|y, u, v_{u,y})}{\sum_{y'} \sum_{u \in U_{y'}} P(zz|y', u, v_{u,y'})},
$$

$$
P(v|zz) = \frac{\sum_{y} \sum_{u \in U_{y,v}} P(zz|y, u, v)}{\sum_{v'} \sum_{y} \sum_{u \in U_{y,v'}} P(zz|y', u, v')}
$$
M step - Binomial on ratings

Factorised:

\[ P(v|z_Z) = \binom{V}{v} p_{zz}^v (1 - p_{zz})^{V-v}, \quad V = |\mathcal{V}|. \]

Binomial with mean \( p_{zz} \cdot V \) and shape parameter \( p_{zz} \).

\[
p_{zz} = \frac{\sum_y \sum_{u \in \mathcal{U}_y} P(z_Z|y,u,v_{u,y}) v_{u,y}}{V \sum_y \sum_{u \in \mathcal{U}_y} P(z_Z|y,u,v_{u,y})}
\]

Type I:

If \( P(v|y,z_Z) \) is binomially distributed, then

\[
p_{zz,y} = \frac{\sum_{u \in \mathcal{U}_y} P(z_Z|y,u,v_{u,y}) v_{u,y}}{V \sum_{u \in \mathcal{U}_y} P(z_Z|y,u,v_{u,y})}
\]
Experiments

Data:
Publicly available *EachMovie dataset* containing ratings ratings by 61,265 users for 1623 films.

User ratings expressed on a 6-point scale from 0.0 to 1.0. The ratings were transformed to \( V = \{1, 2, \ldots, 6\} \).

We selected a set of 100 most rated films. 60,895 users rated at least one film from the selected film set. Final number of ratings: 1,472,253.

Note that the data is still quite sparse. Out of 6,000,895 possible ratings of 100 films by 60,895 users, only 1,472,253 ratings (24.5%) are observed.
Model fitting and validation

Data partitioned into a training set $D$ (train the models and visualize the data) and a test set $T$ (estimate generalization capabilities of the models within the set of users contained in $D$).

All but one protocol: One randomly selected rating from each user having at least 10 ratings was assigned to the test set. The test set contains roughly 3% from all ratings (45,136 ratings).
Performance measures

Normalized negative log likelihood

\[ NNL_{\text{train}} = - \frac{1}{|D|} \sum_{(u, y, v_{uy}) \in D} \log \tilde{P}(v_{uy}, y, u) \]

\[ NNL_{\text{test}} = - \frac{1}{|T|} \sum_{(u, y, v_{uy}) \in T} \log \tilde{P}(v_{uy}, y, u) \]

\[ \tilde{P}(v, y, u) = P(v|y, u) \] and \[ \tilde{P}(v, y, u) = P(v, y|u) \] for models of types I and II, respectively.
Performance measures - Mean absolute deviation

Estimated ratings: $r_{u,y}$
Observed user ratings: $v_{u,y}$

$$
MAD = \frac{1}{|T|} \sum_{(u,y,v_{u,y}) \in T} |r_{u,y} - v_{u,y}|
$$

where $r_{u,y}$ is the most probable rating under the model. For models of type II,

$$
P(v|y, u) = \frac{P(v, y|u)}{\sum_{v \in V} P(v, y|u)}.
$$
Performance measures - Ratio of correctly predicted votes

\[ CPV = \frac{1}{|T|} \sum_{(u,y,v_{u,y}) \in T} E_0(u, y, v_{u,y}), \]

where

\[ E_0(u, y, v_{u,y}) \begin{cases} 
1 & \text{if } r_{u,y} = v_{u,y}, \\
0 & \text{otherwise.} 
\end{cases} \]

A drawback of CPV is that it equally penalizes situations of ‘near miss’ and obviously wrong rating predictions.
Performance measures - Decision accuracy

$T_P$ - set of observations with positive observed ratings, i.e. observations $(u, y, v_{u,y}) \in T$ having $v_{u,y} \geq t = 4$.

$T_R$ - set of observations with positive estimated ratings, i.e. observations $(u, y, v_{u,y}) \in T$ for which $\sum_{r \geq t} P(r|y,u) > \sum_{r < t} P(r|u,y)$.

\[
\text{Precision} = \frac{|T_R \cap T_P|}{|T_R|}
\]

\[
\text{Recall} = \frac{|T_R \cap T_P|}{|T_P|}
\]

\[
\text{F-measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}
\]
## Results - models with grid latent class topology

<table>
<thead>
<tr>
<th></th>
<th>$NNL_{train}$</th>
<th>$NNL_{test}$</th>
<th>MAD</th>
<th>CPV</th>
<th>Precision</th>
<th>Recall</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-Bin</td>
<td>1.39</td>
<td>1.45</td>
<td>1.27</td>
<td>0.31</td>
<td>0.86</td>
<td>0.73</td>
<td>0.79</td>
</tr>
<tr>
<td>I-Mult</td>
<td>1.27</td>
<td>1.42</td>
<td>1.21</td>
<td>0.32</td>
<td>0.86</td>
<td>0.75</td>
<td>0.80</td>
</tr>
<tr>
<td>II-IndB</td>
<td>5.66</td>
<td>5.64</td>
<td>1.40</td>
<td>0.28</td>
<td>0.85</td>
<td>0.70</td>
<td>0.77</td>
</tr>
<tr>
<td>II-IndM</td>
<td>5.56</td>
<td>5.54</td>
<td>1.37</td>
<td>0.29</td>
<td>0.86</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>II-Mult</td>
<td>5.53</td>
<td>5.53</td>
<td>1.31</td>
<td>0.31</td>
<td>0.86</td>
<td>0.68</td>
<td>0.76</td>
</tr>
</tbody>
</table>
## Results - models without latent class topology

<table>
<thead>
<tr>
<th></th>
<th>$NNL_{\text{train}}$</th>
<th>$NNL_{\text{test}}$</th>
<th>MAD</th>
<th>CPV</th>
<th>Precision</th>
<th>Recall</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-Bin</td>
<td>1.25</td>
<td>1.46</td>
<td>1.21</td>
<td>0.32</td>
<td>0.87</td>
<td>0.76</td>
<td>0.81</td>
</tr>
<tr>
<td>I-Mult</td>
<td>0.97</td>
<td>2.37</td>
<td>1.31</td>
<td>0.29</td>
<td>0.85</td>
<td>0.75</td>
<td>0.80</td>
</tr>
<tr>
<td>II-IndB</td>
<td>5.14</td>
<td>7.76</td>
<td>1.76</td>
<td>0.24</td>
<td>0.85</td>
<td>0.56</td>
<td>0.67</td>
</tr>
<tr>
<td>II-IndM</td>
<td>4.94</td>
<td>10.43</td>
<td>1.51</td>
<td>0.26</td>
<td>0.84</td>
<td>0.65</td>
<td>0.73</td>
</tr>
<tr>
<td>II-Mult</td>
<td>4.90</td>
<td>11.25</td>
<td>1.40</td>
<td>0.27</td>
<td>0.85</td>
<td>0.66</td>
<td>0.74</td>
</tr>
</tbody>
</table>
Understanding the models

For each latent class on the grid, we present 5 most probable films, i.e. films with largest $P(y|z) = \sum_v P(v,y|z)$.

In addition to film names, we show genre codes from Internet Movie Database. There are 17 genres which we order into a template $[A,V,N,L,C,P,D,F,Y,H,M,U,R,S,T,W,E]$. E.g.

‘A’ - Action
‘V’ - Adventure
‘N’ - Animation
‘L’ - Classic
‘P’ - Crime
‘C’ - Comedy
‘D’ - Drama

The film genres were not explicitly used in training the models.
Latent Space Modeling in CF

| -----C------S-- | AV-----D-----ST-- | A--C--------S-- | -----D------T-- | A----------ST-- | A-----D------T-- | -----C------S-- |
| A--C----------- | -----C------------ | -----C-FY------ | A-----D------T-- | AV---------H-----ST-- | AV-----Y------S-- | -----C------S-- |
| AV---------H-U-S-- | AV----------T-- | -----D------T-- | A-----D------T-- | AV---------U-----T-- | AV----------T-- | -----C------S-- |
| -----C----------- | -----C---------E | A--D--------T-- | -----D------T-- | A----------T-- | AV--C------R-T-- | -----C------S-- |
| -----C----------- | A--------------T-- | A--CP--Y------ | A--CP--Y------ | -----C-D--------T-- | -----C-FY------ | A--C--------S-- |
| AV----------T-- | A-----D------T-- | A-----D------T-- | -----D------T-- | A----------T-- | -----C------S-- | -----C------S-- |
| AV---------Y------S-- | AV---------Y------S-- | AV----------S-- | -----C-F------R-- | -----C-DF------ | -----C------S-- | -----C------S-- |
| AV---------D-----ST-- | -----D------H-- | A-----D------T-- | A--CP--Y------ | VN------F-----M-- | AV---------H-----ST-- | -----C------S-- |
| AV---------P--Y--R-T-- | AV--CP--------T-- | AV--P--------T-- | A-----D------T-- | A----------ST-- | A-----D------W-- | -----C------S-- |
| A-----D------T-- | AV--P--------T-- | -----V------C-- | -----D------T-- | A--C------R-T-- | A-----D------W-- | -----C------S-- |
| AV---------P--Y--R-T-- | AV--CP--------T-- | A-----D------T-- | -----D------T-- | -----C------R-- | A-----D------W-- | -----C------S-- |
| A-----P--Y------T-- | A-----D------T-- | A-----D------T-- | AV---------H-----ST-- | -----C-D------R-- | -----C------S-- | -----C------S-- |
| AV--CP--------T-- | AV--C------R-T-- | -----N-C--FY-M-- | -----N------F-----M-- | -----D------T-- | AV----------U-----T-- | -----C------S-- |
| -----C----------- | -----N-C--FY-M-- | AV--C------R-T-- | A----------ST-- | -----C-D------Y----R-T-- | -----P--H------T-- | -----C------S-- |
| AV---------Y------S-- | -----N------F-----M-- | AV--CP--------T-- | -----V------N--F--M-- | -----C-DF-------- | A--CPD-------- | -----C------S-- |
| -----C----------- | -----D------T-- | -----D------T-- | AV---------H-----ST-- | -----C------R-- | -----C------S-- | -----C------S-- |
| -----V------C-- | -----PD-------- | -----P--H------T-- | -----C------R-- | -----D------T-- | A-----------S--- | -----C------S-- |
| -----PD-------- | -----V------D------E | -----PD-------- | A----------T-- | A-----D------T-- | -----D------T-- | -----C------S-- |
| AV---------P--Y--R-T-- | -----D------T-- | -----D------T-- | A----------ST-- | -----C-D------Y----R-T-- | -----P--H------T-- | -----C------S-- |
| -----N-C--FY-M-- | -----P--H------T-- | AV----------U-----T-- | -----V------N--F--M-- | -----C------R-- | -----D------T-- | -----C------S-- |
| A-----P--Y------T-- | A-----P--Y------T-- | -----V------D------E | -----C------R-- | -----D------R-- | -----C------R-- | -----C------S-- |
| AV--C------R-T-- | -----D------T-- | -----D------T-- | AV---------H-----ST-- | -----C------R-- | -----C------S-- | -----C------S-- |
| -----V------D------E | -----V------D------E | -----N-C--FY-M-- | AV----------U-----T-- | -----C-DF-------- | A--CPD-------- | -----C------S-- |
| -----PD-------- | -----D------T-- | AV----------U-----T-- | A----------ST-- | -----C------R-- | -----D------T-- | -----C------S-- |
| AV---------P--Y--R-T-- | -----N-C--FY-M-- | -----PD-------- | A----------T-- | -----U------T-- | -----D------T-- | -----C------S-- |
| -----N-C--FY-------- | -----P--H------T-- | AV----------U-----T-- | -----V------N--F--M-- | -----C------R-- | -----D------T-- | -----C------S-- |
| A--C--------S-- | -----D------T-- | -----D------ST-- | -----C------R-- | -----C-DF-------- | A--CPD-------- | -----C------S-- |
| -----D------ST-- | -----D------ST-- | A----------T-- | -----C-D------Y----R-T-- | AV--D------ST-- | -----D--24ST-- | -----C------S-- |
| -----D------ST-- | A--D--------T-- | A--D--------T-- | -----C-D------Y----R-T-- | AV--D------ST-- | -----D--24ST-- | -----C------S-- |

P. Tino and G. Polcicova

Latent Space Modeling in CF
Understanding the models

Characterise the class-conditional film and rating distributions $P(y|z)$ and $P(v|z)$, respectively, via

- modes
- means
- Normalized entropies

\[
H[P(y|z)] = \sum_{y \in Y} P(y|z) \log_2 |Y| P(y|z)
\]

\[
H[P(v|z)] = \sum_{v \in V} P(v|z) \log_2 |V| P(v|z)
\]
Understanding the models

\[ H[P(y|z)] \]

\[ H[P(v|z)] \]

Mean of \( P(v|z) \)

Mode of \( P(v|z) \)
Conclusions

- Topographic versions of two latent class models for collaborative filtering.

- Topographic organization can help to better understand hidden patterns in large and sparse rating databases.

- The preference patterns can be inspected in a variety of ways. Use tools from probability and information theories to interpret and visualize trends captured by the abstract latent classes.

- Several distribution models to account for (class conditional) variations in user ratings.

- Multinomial distribution is adequate if the model is regularized by the tight grid topology on the latent space.

- Possible applications - market research, air-time scheduling etc.
More details in …

P. Tiňo, G. Polčicová:
Topographic organization of user preference patterns in collaborative filtering.

G. Polčicová, P. Tiňo:
Making sense of sparse rating data in collaborative filtering via topographic organization of user preference patterns.