Collaborative Filtering via Group-Structured Dictionary Learning

Zoltán Szabó[1] **Barnabás Póczos**[2] **András Lőrincz**[1]

[1]Faculty of Informatics, Eötvös Loránd University
Pázmány Péter sétány 1/C, Budapest, H-1117, Hungary
email: `{szzoli,andras.lorincz}@elte.hu`
web: `http://nipg.inf.elte.hu`

[2]School of Computer Science, Carnegie Mellon University
5000 Forbes Ave, 15213, Pittsburgh, PA, USA
email: `bapoczos@cs.cmu.edu`
web: `http://www.autonlab.org`

1. Introduction

- Help users in decision making.
- Recommender systems: collaborative filtering (CF) [1]:
  - users' preferences = ratings,
  - estimation based on (i) his/her rating history, (ii) ratings of similar users.
- Novel advances in CF: dictionary learning (latent, unstructured features).
- **Our goal**:
  - structured dictionaries [2] to CF.
  - +requirements:
    - ∗ online learning: changing item-/user set/preferences; adaptation.
    - ∗ incomplete observations: missing rating values.

2. The OSDL Problem

Definition [3]:

- Group structure inducing on the hidden representation $\boldsymbol{\alpha}$ through regularization:

$$\Omega(\boldsymbol{\alpha}) = \|(\|\boldsymbol{\alpha}_G\|_2)_{G\in\mathcal{G}}\|_\eta, \quad \eta \in (0,2). \tag{1}$$

- Approximate on the observed coordinates ($\mathbf{x}_O$) using dictionary $\mathbf{D}$:

$$\frac{1}{2}\|\mathbf{x}_O - \mathbf{D}_O\boldsymbol{\alpha}\|_2^2. \tag{2}$$

- Loss for a fixed observation ($\kappa > 0$):

$$l(\mathbf{x}_O, \mathbf{D}_O) = \min_{\boldsymbol{\alpha}\in\mathcal{A}} \left[\frac{1}{2}\|\mathbf{x}_O - \mathbf{D}_O\boldsymbol{\alpha}\|_2^2 + \kappa\Omega(\boldsymbol{\alpha})\right]. \tag{3}$$

- Goal: minimize online the average loss of the dictionary ($\rho = 0$)

$$\min_{\mathbf{D}\in\mathcal{D}} f_t(\mathbf{D}) := \frac{1}{t}\sum_{i=1}^{t} l(\mathbf{x}_{O_i}, \mathbf{D}_{O_i}). \tag{4}$$

Inclusion of forgetting ($\rho \geq 0$) is possible/motivated:

$$\min_{\mathbf{D}\in\mathcal{D}} f_t(\mathbf{D}) := \frac{1}{\sum_{j=1}^{t}(j/t)^\rho}\sum_{i=1}^{t}\left(\frac{i}{t}\right)^\rho l(\mathbf{x}_{O_i}, \mathbf{D}_{O_i}). \tag{5}$$

Special cases for $\mathcal{G}$:

'Traditional' sparse dictionary	$\mathcal{G} = \{\{1\},\{2\},\ldots,\{d_\alpha\}\}$.
Group Lasso	$\mathcal{G}$ = partition.
Hierarchical dictionary	$\mathcal{G}$ = descendants of the nodes.
Grid adopted dictionary	$\mathcal{G}$ = nearest neighbors of the nodes.

Online optimization of dictionary $\mathbf{D}$ through alternations[1]:

1. Representation update ($\boldsymbol{\alpha}_t$): variational property of $\|\cdot\|_\eta$.
2. Dictionary update ($\mathbf{D}_t$):
   - update statistics of the cost $\hat{f}_t$: matrix recursions.
   - block-coordinate descent optimization.

3. CF Task via OSDL

- t^{th} user's known ratings = OSDL observations $\mathbf{x}_{O_t}$} $\Rightarrow \mathbf{D}$.
- Test user ($\mathbf{x}_O \in \mathbb{R}^{|O|}$):
  1. Estimate $\boldsymbol{\alpha}$: using $\mathbf{x}_O$ and $\mathbf{D}_O$ (rows of $\mathbf{D}$ restricted to O; solve (3)).
  2. Estimate ratings: $\hat{\mathbf{x}} = \mathbf{D}\boldsymbol{\alpha}$.
- Neighbor correction for further improvement:
  - assumption: similar items are rated similarly (s_{ij}).
  - OSDL estimation ($\mathbf{d}_k\boldsymbol{\alpha}_t$, $k \notin O_t$) corrected by the similarity weighted (s_{kj}) prediction errors ($\mathbf{d}_j\boldsymbol{\alpha}_t - x_{jt}$) of the observable items ($j \in O_t$).

4. Numerical Results

- Dataset: joke recommendation (Jester), 100 jokes $\times$ $73,421$ users ($4,136,360$ ratings).
- Performance measure:

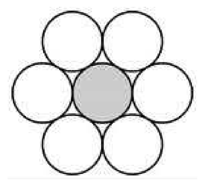
$$RMSE = \sqrt{\frac{1}{|\mathcal{S}|}\sum_{(i,t)\in\mathcal{S}}(x_{it} - \hat{x}_{it})^2}. \tag{6}$$

- Baseline: best known RMSE = 4.1123 (item neighbor), 4.1229 (unstructured dictionary).
- Applied similarities [$s_{ij} = s_{ij}(\mathbf{d}_i, \mathbf{d}_j)$, $\beta > 0$]:

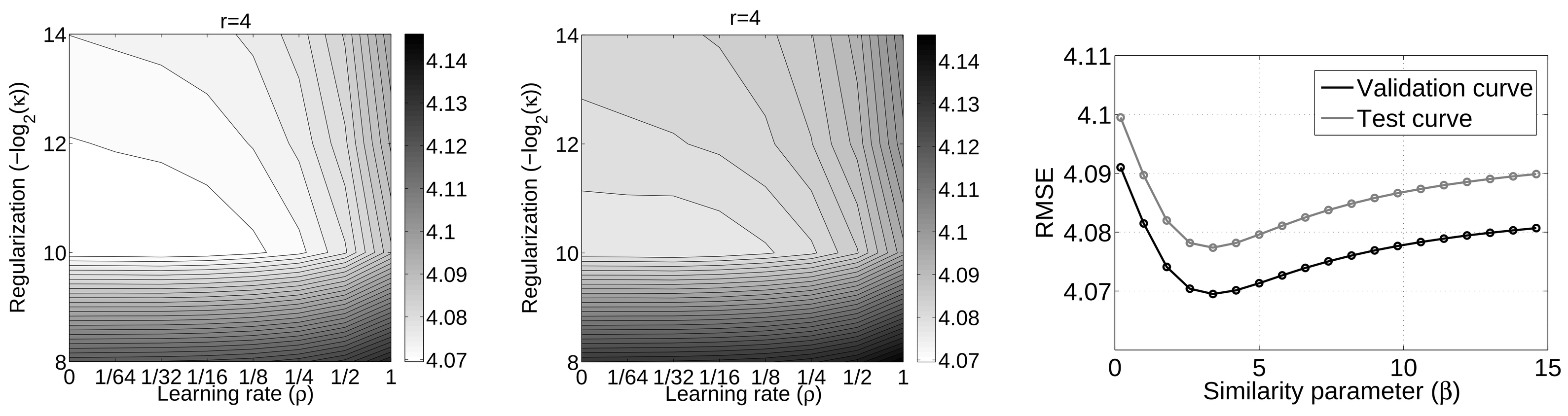
$$S_1: \quad s_{ij} = \left(\frac{\max\left(0, \mathbf{d}_i\mathbf{d}_j^T\right)}{\|\mathbf{d}_i\|_2\|\mathbf{d}_j\|_2}\right)^\beta, \text{ and } S_2: \quad s_{ij} = \left(\frac{\|\mathbf{d}_i - \mathbf{d}_j\|_2^2}{\|\mathbf{d}_i\|_2\|\mathbf{d}_j\|_2}\right)^{-\beta}. \tag{7}$$

$s_{ij} \geq 0$, close to zero (large) = very different (very similar) items.

Toroid group structure: varying neighbor size ($r \in \{0,1,\ldots,5\}$), Fig. 1, Table 2.



- validation, test surfaces: very similar.
- the same holds for similarity parameter (β) dependence.
- structured dictionaries ($r > 0$) are advantageous over unstructured ones ($r = 0$).
- best result ($r = 4$): RMSE $= 4.0774 <$ state-of-the-art (RMSE $= 4.1123$).
- robust estimation w.r.t. forgetting factor (ρ), similarity (S_i) and mini-batch size (R).



(a) (b) (c)

Figure 1: *(a)-(b): validation and test surface – forgetting factor and regularization dependence. (c): validation and test curves – similarity parameter dependence.*

Table 2: *OSDL prediction – performance summary. Group structure ($\mathcal{G}$): toroid.*

	$R=8$					$R=16$				
	$r=0$	$r=1$	$r=2$	$r=3$	$r=4$	$r=0$	$r=1$	$r=2$	$r=3$	$r=4$
S_1	4.1594	4.1326	4.1274	4.0792	**4.0774**	4.1611	4.1321	4.1255	4.0804	**4.0777**
S_2	4.1765	4.1496	4.1374	4.0815	4.0802	4.1797	4.1487	4.1367	4.0826	4.0802

Hierarchical group structure: varying hierarchy level ($l \in \{3,\ldots,6\}$).

- Results: similar to that of the toroid structure.
- Best RMSE = 4.1220 ($l = 4$, i.e., $d_\alpha = 15$).
  - much smaller d_α compared to unstructured dictionaries ($d_\alpha = 100$, RMSE$= 4.1229$).
  - competitive to the state-of-the-art (RMSE$= 4.1123$).

References

[1] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul Kantor. *Recommender Systems Handbook*. Springer, 2011.

[2] Francis Bach, Rodolphe Jenatton, Julien Mairal, and Guillaume Obozinski. *Optimization for Machine Learning*, chapter Convex optimization with sparsity-inducing norms. MIT Press, 2011.

[3] Zoltán Szabó, Barnabás Póczos, and András Lőrincz. Online group-structured dictionary learning. In *CVPR 2011*, pages 2865–2872.

The research was partly supported by the Department of Energy (grant number DESC0002607).

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

The project is supported by the European Union and co-financed by the European Social Fund (grant no. TÁMOP 4.2.1/B-09/1/KMR-2010-0003) and the European Regional Development Fund (grant no. KMOP-1.1.2-08/1-2008-0002).

[1]Matlab code available at `http://nipg.inf.elte.hu/szzoli`.