NCDREC: A Decomposability Inspired Framework for Top-N Recommendation

Athanasios N. Nikolakopoulos[1,2] John D. Garofalakis[1,2]

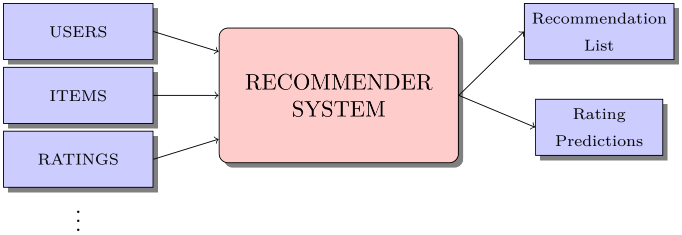
[1]Computer Engineering and Informatics Department,
University of Patras, Greece

[2]Computer Technology Institute & Press "Diophantus"

IEEE/WIC/ACM International Conference on Web Intelligence
Warsaw, August 2014

Outline

Recommender System Algorithms



- **Collaborative Filtering Recommendation Algorithms**
  - Wide deployment in Commercial Environments
  - Significant Research Efforts

Challenges of Modern CF Algorithms

Sparsity

- It is an Intrinsic RS Characteristic related to serious problems:
  - **Long-Tail Recommendation**
  - **Cold start Problem**
  - **Limited ItemSpace Coverage**

Traditional CF techniques, such as neighborhood models, are very susceptible to sparsity. Among the most promising approaches in alleviating sparsity related problems are:

- **Dimensionality-Reduction Models**
  - Build a reduced latent space which is dense.
- **Graph-Based Models**.
  - Exploit transitive relations in the data, while preserving some of the "locality".

Exploiting Decomposability

We attack the problem from a different perspective:

- Sparsity $\longleftrightarrow$ Hierarchy $\longleftrightarrow$ Decomposability.
- **Nearly Completely Decomposable Systems**
  - Pioneered by **Herbert A. Simon**.
  - Many Applications in Diverse Disciplines such as economics, cognitive theory and social sciences, to computer systems performance evaluation, data mining and information retrieval
- **Main Idea**: Exploit the innate Hierarchy of the Item Set, and view it as a decomposable space.
  - Can this enrich the Collaborative Filtering Paradigm in an Efficient and Scalable Way?
  - Does this approach offer any qualitative advantages in alleviating sparsity related problems?

NCDREC Model Overview

Definitions

- We define a $\mathcal{D}$-**decomposition** to be an indexed family of sets $\mathcal{D} \triangleq \{\mathcal{D}_1, \dots, \mathcal{D}_K\}$, that span the ItemSpace $\mathcal{V}$,
- We define $\mathfrak{D}_v \triangleq \bigcup_{v \in \mathcal{D}_k} \mathcal{D}_k$ to be the **proximal set** of items of $v \in \mathcal{V}$,
- We also define the associated **block coupling graph** $\mathcal{G}_\mathcal{D} \triangleq (\mathcal{V}_\mathcal{D}, \mathcal{E}_\mathcal{D})$; its vertices correspond to the $\mathcal{D}$-blocks, and an edge between two vertices exists whenever the intersection of these blocks is a non-empty set. Finally, we introduce an **aggregation matrix** $\mathbf{A}_\mathcal{D} \in \mathbb{R}^{m \times K}$, whose jk^{th} element is 1, if $v_j \in \mathcal{D}_k$ and zero otherwise.

NCDREC Components

- **Main Component: Recommendation vectors produced by projecting the NCD perturbed data onto an f-dimensional space.**
- **ColdStart Component:Recommendation vectors are the stationary distributions of a Discrete Markov Chain Model.**

$$\mathbf{G} \triangleq \mathbf{R} + \epsilon \mathbf{W}$$

$$\mathbf{W} \triangleq \epsilon \mathbf{Z}\mathbf{X}^\intercal$$

$$\mathbf{X} \triangleq \operatorname{diag}(\mathbf{A}_\mathcal{D}\mathbf{e})^{-1}\mathbf{A}_\mathcal{D}$$

$$[\mathbf{Z}]_{ik} \triangleq (n^k_{u_i})^{-1}[\mathbf{R}\mathbf{A}_\mathcal{D}]_{ik}, \text{ when } n^k_{u_i} > 0,$$

and zero otherwise.

$$\mathbf{S}(\boldsymbol{\omega}) \triangleq (1-\alpha)\mathbf{E} + \alpha(\beta\mathbf{H} + (1-\beta)\mathbf{D})$$

$$\mathbf{H} \triangleq \operatorname{diag}(\mathbf{C}\mathbf{e})^{-1}\mathbf{C},$$

where $[\mathbf{C}]_{ij} \triangleq \mathbf{r}_\mathbf{i}^\intercal \mathbf{r}_\mathbf{j}$ for $i \neq j$

$$\mathbf{D} \triangleq \mathbf{X}\mathbf{Y}, \quad \mathbf{X} \triangleq \operatorname{diag}(\mathbf{A}_\mathcal{D}\mathbf{e})^{-1}\mathbf{A}_\mathcal{D},$$

$$\mathbf{Y} \triangleq \operatorname{diag}(\mathbf{A}_\mathcal{D}^\intercal\mathbf{e})^{-1}\mathbf{A}_\mathcal{D}^\intercal,$$

$$\mathbf{E} \triangleq \mathbf{e}\boldsymbol{\omega}^\intercal$$

Criterion for ItemSpace Coverage

Theorem (ItemSpace Coverage)

If the block coupling graph $\mathcal{G}_{\mathcal{D}}$ is connected, there exists a unique steady state distribution $\boldsymbol{\pi}$ of the Markov chain corresponding to matrix $\mathbf{S}$ that depends on the preference vector $\boldsymbol{\omega}$; however, irrespectively of any particular such vector, the support of this distribution includes every item of the underlying space.

Proof Sketch: When $\mathcal{G}_{\mathcal{D}}$ is connected, the Markov chain induced by the stochastic matrix $\mathbf{S}$ consists of a single irreducible and aperiodic closed set of states, that includes all the items. The above is true for every stochastic vector $\boldsymbol{\omega}$, and for every positive real numbers $\alpha, \beta < 1$. Taking into account the fact that the state space is finite, the resulting Markov chain becomes ergodic. So $\pi_i > 0$, for all i, and the support of the distribution that defines the recommendation vector includes every item of the underlying space.

NCDREC Algorithm: Storage and Computational Issues

Input: Matrices $\mathbf{R} \in \mathbb{R}^{n \times m}$, $\mathbf{H} \in \mathbb{R}^{m \times m}$, $\mathbf{X} \in \mathbb{R}^{m \times K}$, $\mathbf{Y} \in \mathbb{R}^{K \times m}$, $\mathbf{Z} \in \mathbb{R}^{n \times K}$. Parameters $\alpha, \beta, f, \epsilon$
Output: The matrix with recommendation vectors for every user, $\mathbf{\Pi} \in \mathbb{R}^{n \times m}$
Step 1: Find the newly added users and collect their preference vectors into matrix $\mathbf{\Omega}$.
Step 2: Compute $\mathbf{\Pi}_{\text{sparse}}$ using the ColdStart Procedure.
Step 3: Initialize vector $\mathbf{p_1}$ to be a random unit length vector.
Step 4: Compute the modified Lanczos procedure up to step M, using NCD_PartialLBD with starting vector $\mathbf{p_1}$.
Step 5: Compute the SVD of the bidiagonal matrix $\mathbf{B}$ and use it to extract $f < M$ approximate singular triplets:

$$\{\tilde{\mathbf{u}}_{\mathbf{j}}, \sigma_j, \tilde{\mathbf{v}}_{\mathbf{j}}\} \leftarrow \{\mathbf{Q}\mathbf{u}_{\mathbf{j}}^{(\mathbf{B})}, \sigma_j^{(B)}, \mathbf{P}\mathbf{v}_{\mathbf{j}}^{(\mathbf{B})}\}$$

Step 6: Orthogonalize against the approximate singular vectors to get a new starting vector $\mathbf{p_1}$.
Step 7: Continue the Lanczos procedure for M more steps using the new starting vector.
Step 8: Check for convergence tolerance. If met compute matrix:

$$\mathbf{\Pi}_{\text{full}} = \tilde{\mathbf{U}}\mathbf{\Sigma}\tilde{\mathbf{V}}^{\intercal}$$

else go to **Step 4**
Step 9: Update $\mathbf{\Pi}_{\text{full}}$, replacing the rows that correspond to new users with $\mathbf{\Pi}_{\text{sparse}}$.
Return $\mathbf{\Pi}_{\text{full}}$

Experimental Evaluation

Datasets

- Yahoo!R2Music
- MovieLens

Competing Methods

- *Commute Time* (CT)
- *Pseudo-Inverse of the user-item graph Laplacian* ($L^\dagger$)
- *Matrix Forest Algorithm* (MFA)
- *First Passage Time* (FP)
- *Katz Algorithm* (Katz)

Metrics

- **Recall**
- **Precision**
- **R-Score**
- **Normalized Discounted Cumulative Gain (NDCG@k)**
- **Mean Reciprocal Rank**

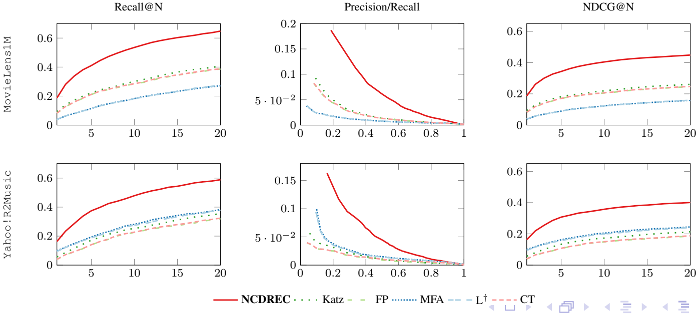
Full Dist. Recommendations

Methodology

- Randomly sample 1.4% of the ratings of the dataset ⇒ probe set $\mathcal{P}$
- Use each item v_j, rated with 5 stars by user u_i in $\mathcal{P}$ ⇒ test set $\mathcal{T}$
- Randomly select another 1000 unrated items of the same user for each item in $\mathcal{T}$
- Form ranked lists by ordering all the 1001 items

TABLE I
RECOMMENDATION QUALITY ON MOVIELENS1M AND YAHOO!R2MUSIC DATASETS USING R-SCORE AND MRR METRICS

	MovieLens1M			Yahoo!R2Music		
	R(5)	R(10)	MRR	R(5)	R(10)	MRR
NCDREC	**0.3997**	**0.5098**	**0.3008**	**0.3539**	**0.4587**	**0.2647**
MFA	0.1217	0.1911	0.0887	0.2017	0.2875	0.1591
$L^{\dagger}$	0.1216	0.1914	0.0892	0.1965	0.2814	0.1546
FP	0.2054	0.2874	0.1524	0.1446	0.2241	0.0998
Katz	0.2187	0.3020	0.1642	0.1704	0.2529	0.1203
CT	0.2070	0.2896	0.1535	0.1465	0.2293	0.1019



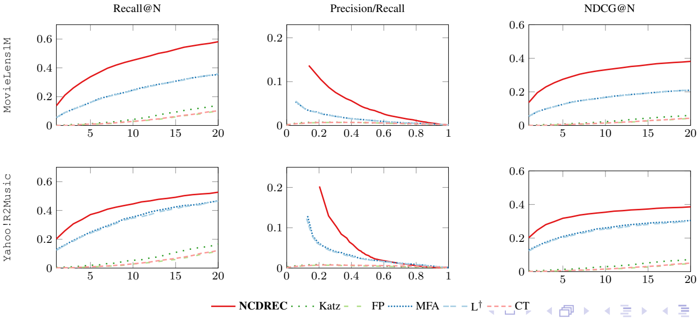
Long-Tail Recommendations

Methodology (Long Tail)

- We order the items according to their popularity (measured in terms of number of ratings)
- We further partition the test set $\mathcal{T}$ into two subsets, $\mathcal{T}_{\text{head}}$ and $\mathcal{T}_{\text{tail}}$
- We discard the popular items and we evaluate NCDREC and the other algorithms on the $\mathcal{T}_{\text{tail}}$ test set.

TABLE II
LONG TAIL RECOMMENDATION QUALITY ON MOVIELENS1M AND YAHOO!R2MUSIC DATASETS USING R-SCORE AND MRR METRICS

	MovieLens1M			Yahoo!R2Music		
	R(5)	R(10)	MRR	R(5)	R(10)	MRR
NCDREC	**0.3279**	**0.4376**	**0.2395**	**0.3520**	**0.4322**	**0.2834**
MFA	0.1660	0.2517	0.1188	0.2556	0.3530	0.1995
$L\dagger$	0.1654	0.2507	0.1193	0.2492	0.3461	0.1939
FP	0.0183	0.0654	0.0221	0.0195	0.0684	0.0224
Katz	0.0275	0.0822	0.0267	0.0349	0.0939	0.0309
CT	0.0192	0.0675	0.0227	0.0215	0.0747	0.0249



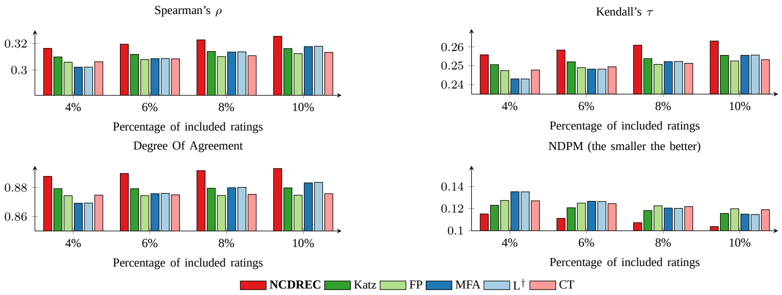
New Users problem

Methodology

- Randomly select 100 users having rated at least 100 items and delete 96%, 94%, 92% and 90% of each users' ratings.
- Compare the rankings induced on the modified data with the complete set of ratings.

Metrics

- **Spearman's ρ**
- **Kendall's τ**
- **Degree of Agreement** (DOA)
- **Normalized Distance-based Performance Measure** (NDPM)



Future Research Directions & Conclusion

Future Work

- Decomposition Granularity Effect
  - Coarse Grained $\Rightarrow$ Sparseness Insensitivity
  - Fine Grained $\Rightarrow$ Higher Quality of Recommendations
- Multiple-Criteria Decompositions
  - How it affects the theoretical properties of the ColdStart Component?

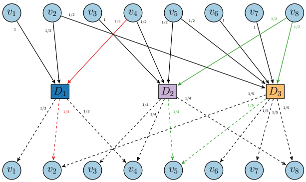
Thanks!

Q&A

Example NCD Proximity Matrix

Back

$$
\begin{array}{c}
\\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8
\end{array}
\begin{array}{c}
\begin{array}{ccccc} \mathcal{D}_1 & \mathcal{D}_2 & \mathcal{D}_3 & \mathcal{N}_v & \mathcal{G}_v \end{array} \\
\left(\begin{array}{ccc||c|c}
\checkmark & - & - & 1 & \{v_1, v_2, v_4\} \\
\checkmark & - & \checkmark & 2 & \{v_1, v_2, v_4, v_5, v_6, v_7, v_8\} \\
- & \checkmark & - & 1 & \{v_3, v_4, v_5, v_8\} \\
\checkmark & \checkmark & - & 2 & \{v_1, v_2, v_3, v_4, v_5, v_8\} \\
- & \checkmark & \checkmark & 2 & \{v_2, v_3, v_4, v_5, v_6, v_7, v_8\} \\
- & - & \checkmark & 1 & \{v_2, v_5, v_6, v_7, v_8\} \\
- & - & \checkmark & 1 & \{v_2, v_5, v_6, v_7, v_8\} \\
- & \checkmark & \checkmark & 2 & \{v_2, v_3, v_4, v_5, v_6, v_7, v_8\}
\end{array}\right)
\end{array}
$$



$\Longrightarrow$

$$
\mathbf{D} = \begin{pmatrix}
\frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\
\frac{1}{6} & \frac{4}{15} & 0 & \frac{1}{6} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\
0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\
\frac{1}{6} & \frac{1}{6} & \frac{1}{8} & \frac{7}{24} & \frac{1}{8} & 0 & 0 & \frac{1}{8} \\
0 & \frac{1}{10} & \frac{1}{8} & \frac{1}{8} & \frac{9}{40} & \frac{1}{10} & \frac{1}{10} & \frac{9}{40} \\
0 & \frac{1}{5} & 0 & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\
0 & \frac{1}{5} & 0 & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\
0 & \frac{1}{10} & \frac{1}{8} & \frac{1}{8} & \frac{9}{40} & \frac{1}{10} & \frac{1}{10} & \frac{9}{40}
\end{pmatrix}
$$

NCD_PartialLBD

Back

procedure $\text{NCD\_PARTIALLBD}(\mathbf{R}, \mathbf{X}, \mathbf{Z}, \mathbf{p_1}, \epsilon)$

  $\phi \leftarrow \mathbf{X}^\intercal \mathbf{p_1}$; $\mathbf{q_1} \leftarrow \mathbf{R}\mathbf{p_1} + \epsilon \mathbf{Z}\phi$;

  $b_{1,1} \leftarrow \|\mathbf{q_1}\|_2$; $\mathbf{u_1} \leftarrow \mathbf{q_1}/b_{1,1}$;

  for $j = 1$ to M **do**

    $\phi \leftarrow \mathbf{Z}^\intercal \mathbf{q_j}$;

    $\mathbf{r} \leftarrow \mathbf{R}^\intercal \mathbf{q_j} + \epsilon \mathbf{X}\phi - b_{j,j}\mathbf{p_j}$;

    $\mathbf{r} \leftarrow \mathbf{r} - [\mathbf{p_1} \dots \mathbf{p_j}]\left([\mathbf{p_1} \dots \mathbf{p_j}]^\intercal \mathbf{r}\right)$;

    if $j < M$ **then**

      $b_{j,j+1} \leftarrow \|\mathbf{r}\|$; $\mathbf{p_{j+1}} \leftarrow \mathbf{r}/b_{j,j+1}$;

      $\phi \leftarrow \mathbf{X}^\intercal \mathbf{p_{j+1}}$;

      $\mathbf{q_{j+1}} \leftarrow \mathbf{R}\mathbf{p_{j+1}} + \epsilon \mathbf{Z}\phi - b_{j,j+1}\mathbf{q_j}$;

      $\mathbf{q_{j+1}} \leftarrow \mathbf{q_{j+1}} - [\mathbf{q_1} \dots \mathbf{q_j}]\left([\mathbf{q_1} \dots \mathbf{q_j}]^\intercal \mathbf{q_{j+1}}\right)$;

      $b_{j+1,j+1} \leftarrow \|\mathbf{q_{j+1}}\|$;

      $\mathbf{q_{j+1}} \leftarrow \mathbf{q_{j+1}}/b_{j+1,j+1}$;

    end if

  end for

end procedure

ColdStart Procedure

Back

procedure $\textsc{ColdStart}(\mathbf{H}, \mathbf{X}, \mathbf{Y}, \mathbf{\Omega}, \alpha, \beta)$

$\quad \mathbf{\Pi} \leftarrow \mathbf{\Omega}$; $k \leftarrow 0$; $r \leftarrow 1$;

$\quad$ **while** $r > \mathrm{tol}$ and $k \leq \mathrm{maxit}$ **do**

$\quad\quad k \leftarrow k + 1$;

$\quad\quad \hat{\mathbf{\Pi}} \leftarrow \alpha\beta\mathbf{\Pi}\mathbf{H}$; $\mathbf{\Phi} \leftarrow \mathbf{\Pi}\mathbf{X}$;

$\quad\quad \hat{\mathbf{\Pi}} \leftarrow \hat{\mathbf{\Pi}} + \alpha(1-\beta)\mathbf{\Phi}\mathbf{Y} + (1-\alpha)\mathbf{\Omega}$;

$\quad\quad r \leftarrow \|\hat{\mathbf{\Pi}} - \mathbf{\Pi}\|$; $\mathbf{\Pi} \leftarrow \hat{\mathbf{\Pi}}$;

$\quad$ **end while**

return $\mathbf{\Pi}_{\text{sparse}} \leftarrow \mathbf{\Pi}$

end procedure