Boolean Matrix Factorisation for Collaborative Filtering: An FCA-based approach

Dmitry Ignatov[1], Elena Nenova[2], Andrey Konstantinov[1], Natalia Konstantinova[3]

[1]National Research University Higher School of Economics, Moscow, Russia

[2]Imhonet Research, Moscow, Russia

[3]University of Wolverhampton, UK

AIMSA 2014, Sept. 12,
Varna, Bulgaria

Outline

- Problem Statement
- Basic Matrix Factorisation (MF) Techniques
- FCA-based Boolean Matrix Factorisation
  - FCA definitions
  - FCA and Recommender Systems
  - FCA-based BMF
- General Scheme of Experiments
- Experiments
- Conclusion & Future Plans

Problem Statement

- Recommender Systems is a rapidly growing area (ACM RecSys conference series since 2007)
- Matrix Factorisation techniques are seems to be an industry standard (SVD, NMF, PLSA etc.)
- What about Boolean Matrix Factorisation or/and FCA?
- Hence why not to develop FCA-based BMF technique, evaluate it, and compare with the state-of-the-art techniques?

Outline

- Problem Statement
- Basic Matrix Factorisation (MF) Techniques
- FCA-based Boolean Matrix Factorisation
  - FCA definitions
  - FCA and Recommender Systems
  - FCA-based BMF
- General Scheme of Experiments
- Experiments
- Conclusion & Future Plans

Basic MF Techniques. SVD

- Singular Value Decomposition

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T,$$

$A \in \mathbb{R}^{m \times n} (m > n)$,

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices

$\Sigma = diag(\sigma_1, \ldots, \sigma_n)$, where $\sigma_1 \geq \sigma_2 \geq \ldots \geq 0.$

SVD Example

$$A = \begin{pmatrix} 4 & 4 & 5 & 0 & 0 & 0 & 0 \\ 5 & 5 & 3 & 4 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 & 0 & 0 \\ 0 & 0 & 0 & 5 & 4 & 5 & 3 \\ 0 & 0 & 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 0 & 0 & 4 & 4 \end{pmatrix}.$$

$$U = \begin{pmatrix} 0.31 & 0.48 & -0.49 & -0.64 & -0.06 & 0 \\ 0.58 & 0.50 & 0.03 & 0.63 & 0.06 & 0 \\ 0.29 & 0 & 0.57 & -0.23 & -0.72 & 0 \\ 0.57 & -0.37 & 0.31 & -0.30 & 0.57 & 0 \\ 0.29 & -0.47 & -0.43 & 0.15 & -0.28 & -0.62 \\ 0.23 & -0.37 & -0.35 & 0.12 & -0.22 & 0.78 \end{pmatrix},$$

$$\begin{pmatrix} \Sigma \\ 0 \end{pmatrix} = \begin{pmatrix} 12.62 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10.66 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 7.29 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.95 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$V^T = \begin{pmatrix} 0.32 & 0.41 & -0.24 & 0.36 & 0.07 & 0.70 & 0.13 \\ 0.32 & 0.41 & -0.24 & 0.36 & 0.07 & -0.62 & -0.35 \\ 0.26 & 0.37 & -0.32 & -0.79 & -0.12 & -0.06 & 0.17 \\ 0.50 & 0.01 & 0.55 & 0.05 & 0.24 & -0.21 & 0.57 \\ 0.41 & 0.01 & 0.50 & -0.14 & -0.42 & 0.21 & -0.57 \\ 0.42 & -0.53 & -0.27 & -0.15 & 0.57 & 0.10 & -0.28 \\ 0.33 & -0.46 & -0.36 & 0.21 & -0.63 & -0.10 & 0.28 \end{pmatrix}.$$

Basic MF Techniques. NMF

- Non-negative Matrix Factorisation

$$V \approx WH$$

$$V \in \mathbb{R}^{n \times m}, \quad V_{ij} \geq 0;$$

$$W \in \mathbb{R}^{n \times k}, \quad W_{ij} \geq 0;$$

$$H \in \mathbb{R}^{k \times m}, \quad H_{ij} \geq 0.$$

Basic MF Techniques. NMF

$$V = \begin{pmatrix} 4 & 4 & 5 & 0 & 0 & 0 & 0 \\ 5 & 5 & 3 & 4 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 & 0 & 0 \\ 0 & 0 & 0 & 5 & 4 & 5 & 3 \\ 0 & 0 & 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 0 & 0 & 4 & 4 \end{pmatrix}.$$

$$V = \begin{pmatrix} 2.34 & 0 & 0 \\ 2.32 & 1.11 & 0 \\ 0 & 1.28 & 0 \\ 0 & 1.46 & 1.23 \\ 0 & 0 & 1.60 \\ 0 & 0 & 1.28 \end{pmatrix} * \begin{pmatrix} 1.89 & 1.89 & 1.71 & 0.06 & 0 & 0 & 0 \\ 0.13 & 0.13 & 0 & 3.31 & 2.84 & 0.27 & 0 \\ 0 & 0 & 0 & 0.03 & 0 & 3.27 & 2.93 \end{pmatrix}.$$

Basic MF Techniques. NMF

- Boolean Matrix Factorisation

$$I = P \circ Q,$$

$$(P \circ Q)_{ij} = \bigvee_{l=1}^{k} P_{il} \cdot Q_{lj},$$

$$I \in \{0,1\}^{n \times m},$$

$$P \in \{0,1\}^{n \times k},$$

$$Q \in \{0,1\}^{k \times m}.$$

Outline

- Problem Statement
- Basic Matrix Factorisation (MF) Techniques
- **FCA-based Boolean Matrix Factorisation**
  - FCA definitions
  - FCA and Recommender Systems
  - FCA-based BMF
- General Scheme of Experiments
- Experiments
- Conclusion & Future Plans

Formal Concept Analysis

[Wille, 1982, Ganter & Wille, 1999]

Definition 1. Formal Context is a triple (G, M, I), where G is a set of **(formal) objects**, M is a set of **(formal) attributes**, and $I \subseteq G \times M$ is the incidence relation which shows that object $g \in G$ posseses an attribute $m \in M$.

Example. Books recommender

	Romeo & Juliet	**The Puppets Master**	**Ubik**	**Ivanhoe**
Kate	x			x
Mike	x		x	
Alex		x	x	
David		x	x	x

Formal Concept Analysis

Definition 2. Derivation operators (defining Galois connection)

$A^I := \{ m \in M \mid gIm \text{ for all } g \in A \}$ is the set of attributes common to all objects in A

$B^I := \{ g \in G \mid gIm \text{ for all } m \in B \}$ is the set of objects that have all attributes from B

Example

	R&J	**PM**	**Ub**	**Iv**
Kate	x			x
Mike	x		x	
Alex		x	x	
David		x	x	x

$\{Kate, Mike\}^I = \{\mathrm{RJ}\}$

$\{\mathrm{Ubik}\}^I = \{Mike, Alex, David\}$

$\{RJ, PM\}^I = \{\}_{\mathrm{G}}$

$\{\}^I_G = M$

Formal Concept Analysis

Definition 3. (A, B) is *a* ***formal concept*** *of* (G, M, I) iff

$$A \subseteq G,\ B \subseteq M,\ A^I = B, \text{ and } B^I = A\ .$$

A is the **extent** and B is the **intent** of the concept (A, B).

$\mathfrak{B}(G,M,I)$ is a set of all concepts of the context (G, M, I)

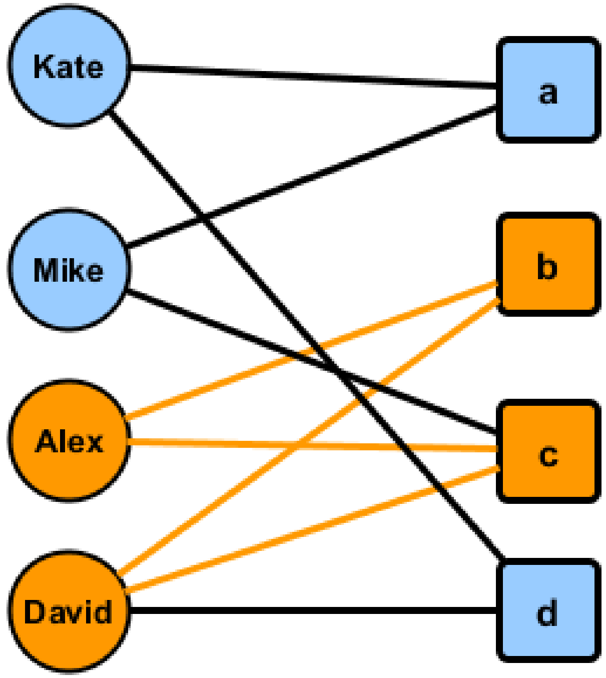
Example

	R&J	**PM**	**Ub**	**Iv**
Kate	x			x
Mike	x		x	
Alex		x	x	
David		x	x	x

- *A pair ({Kate, Mike},*{R&J}) is a formal concept
- *({Alex, David} ,{Ubik}) doesn't form a formal concept, because {Ubik}$^I \neq${Alex, David}*
- *({Alex, David} {PM, Ubik}) is a formal concept*

FCA and Graphs

	a	b	c	d
Kate	x			x
Mike	x		x	
Alex		x	x	
David		x	x	x



Formal Context	Bipartite graph
Formal Concept (maximal rectangle)	Biclique

FCA & Recommender Systems

- Collaborative Recommending using Formal Concept Analysis (du Boucher-Ryan & Bridge, 2006)
- Concept-based Recommendations for Internet Advertisement (Ignatov & Kuznetsov, 2008)
- FCA-based Recommender Models and Data Analysis for Crowdsourcing Platform Witology (Ignatov et al., 2014)

FCA-based BMF

- Belohlavek & Vyhodil, 2010

Matrix I can be considered a matrix of binary relations between set X of objects (users), and a set Y of attributes (items that users have evaluated). We assume that xIy iff the user x evaluated object y. The triple (X, Y, I) clearly forms a formal context.

Consider a set $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$, a subset of all formal concepts of context (X, Y, I), and introduce matrices $P_{\mathcal{F}}$ and $Q_{\mathcal{F}}$:

$$(P_{\mathcal{F}})_{il} = \begin{cases} 1, i \in A_l, \\ 0, i \notin A_l, \end{cases} \quad (Q_{\mathcal{F}})_{lj} = \begin{cases} 1, j \in B_l, \\ 0, j \notin B_l. \end{cases},$$

where (A_l, B_l) is a formal concept from F.

FCA-based BMF

- Belohlavek & Vyhodil, 2010

Theorem 1. (Universality of formal concepts as factors). For every I there is $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$, such that $I = P_{\mathcal{F}} \circ Q_{\mathcal{F}}$.

Theorem 2. (Optimality of formal concepts as factors). Let $I = P \circ Q$ for $n \times k$ and $k \times m$ binary matrices P and Q. Then there exists a $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$ of formal concepts of I such that $|\mathcal{F}| \leq k$ and for the $n \times |\mathcal{F}|$ and $|\mathcal{F}| \times m$ binary matrices $P_{\mathcal{F}}$ and $Q_{\mathcal{F}}$ we have $I = P_{\mathcal{F}} \circ Q_{\mathcal{F}}$.

Example 1

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Example 2

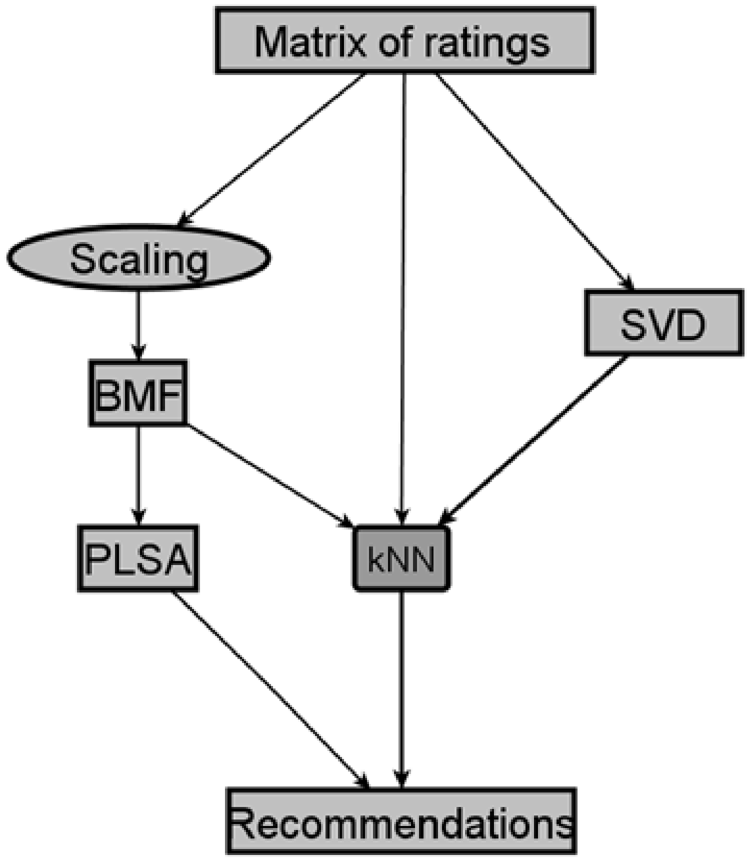
$$\begin{pmatrix} 4 & 4 & 5 & 0 & 0 & 0 & 0 \\ 5 & 5 & 3 & 4 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 & 0 & 0 \\ 0 & 0 & 0 & 5 & 4 & 5 & 3 \\ 0 & 0 & 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 0 & 0 & 4 & 4 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} = I.$$

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Outline

- Problem Statement
- Basic Matrix Factorisation (MF) Techniques
- FCA-based Boolean Matrix Factorisation
  - FCA definitions
  - FCA and Recommender Systems
  - FCA-based BMF
- **General Scheme of Experiments**
- Experiments
- Conclusion & Future Plans

General Scheme of Experiments



kNN approach

- Adomavicus & Tuzhilin, 2005
- Predicted rating of user *c* for item *s*

$$r_{c,s} = k \sum_{c' \in \widehat{C}} sim(c', c) \times r_{c',s},$$

where k serves as a normalizing factor and selected as $k = 1/ \sum_{c' \in \widehat{C}} sim(c, c')$.

- *sim*(*c'*,*c*) is similarity between users *c'* and *c*, e.g. cosine-based or Pearson correlation

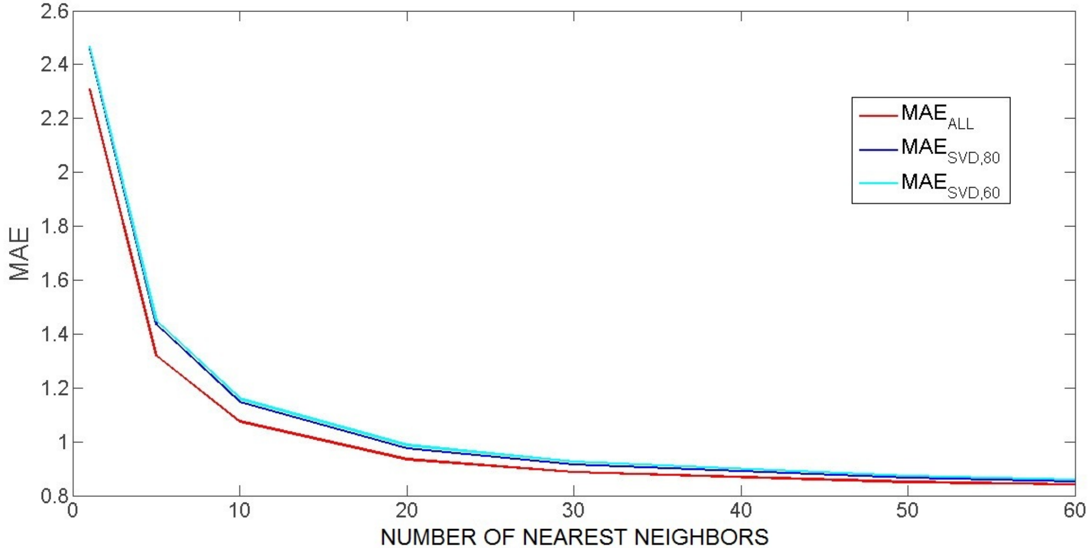
Outline

- Problem Statement
- Basic Matrix Factorisation (MF) Techniques
- FCA-based Boolean Matrix Factorisation
  - FCA definitions
  - FCA and Recommender Systems
  - FCA-based BMF
- General Scheme of Experiments
- Experiments
- Conclusion & Future Plans

Dataset

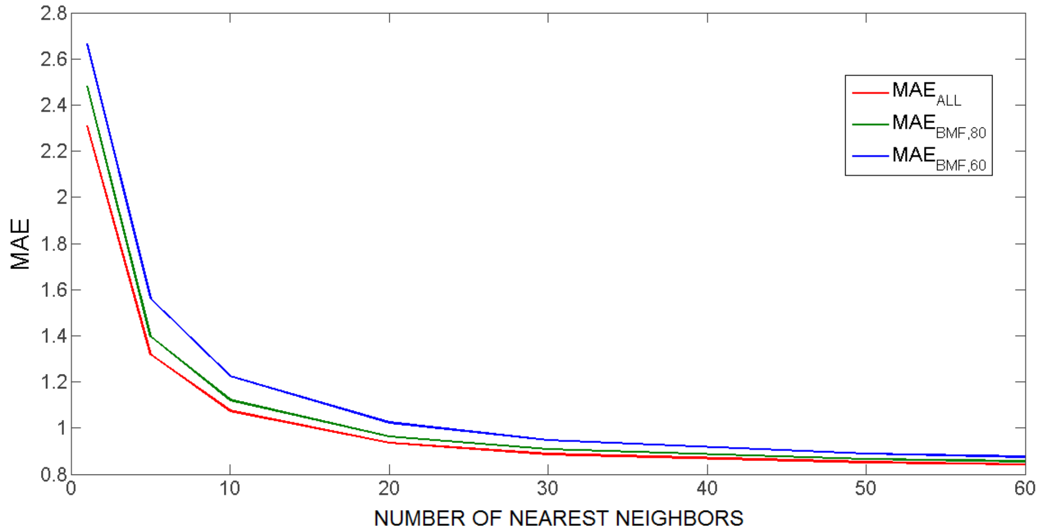
- MovieLens dataset:
  - 943 users,
  - 1682 movies,
  - every user have rated at least 20 movies,
  - 100000 ratings,
  - training set 80000 ratings,
  - test set 20000 ratings.

Experiments



$$p\% = \frac{\sum_{i=1}^{K} \sigma_i^2}{\sum \sigma_i^2} * 100\%,$$

$$p\% = \frac{used\_marks}{all\_marks} * 100\%.$$



Experiments

- MAE for SVD and BMF at 80% coverage level

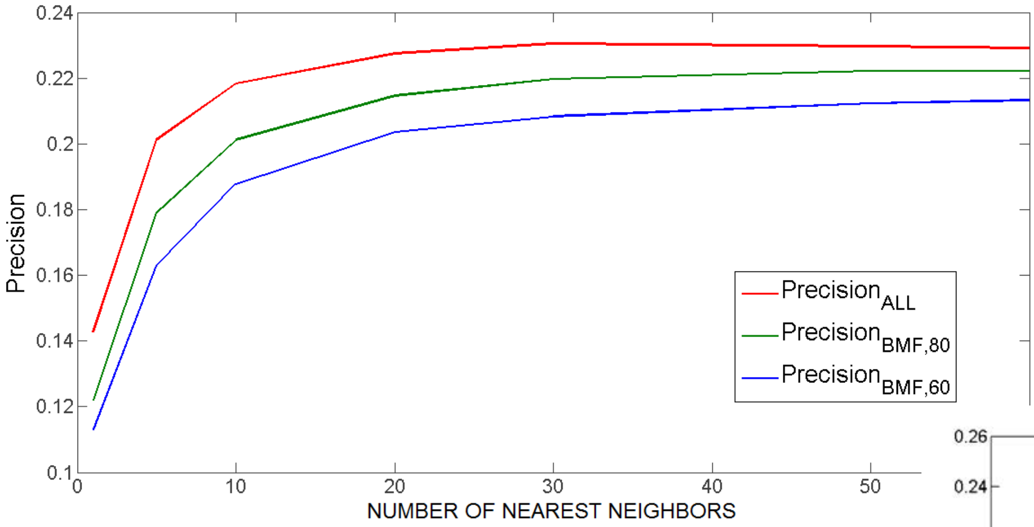
Number of neighbors	1	5	10	20	30	50	60
MAE_{SVD80}	**2,4604**	1.4355	1.1479	0.9750	0.9148	0.8652	**0.8534**
MAE_{BMF80}	2.4813	**1.3960**	**1.1215**	**0.9624**	**0.9093**	**0.8650**	0.8552
MAE_{all}	2.3091	1.3185	1.0744	0.9350	0.8864	0.8509	0.8410

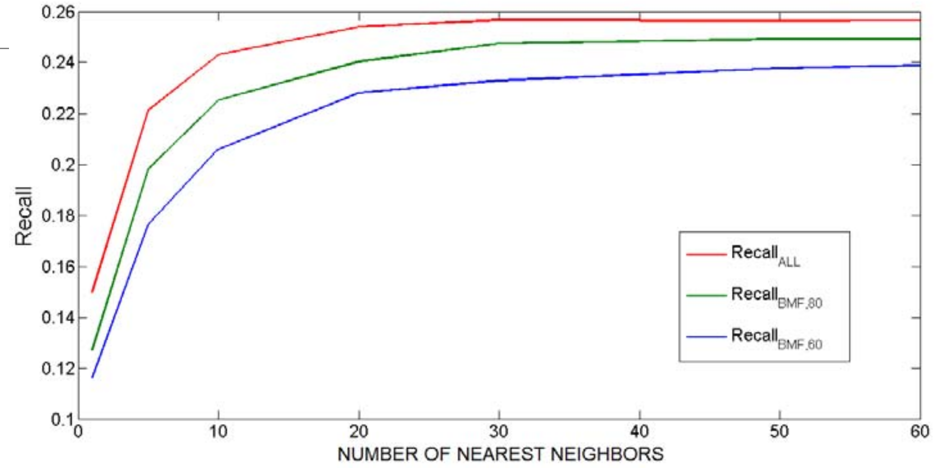
- Number of factors for SVD and BMF at different coverage level

p%	100%	80%	60%
SVD	943	175	67
BMF	1302	402	223

Experiments

- Comparison of kNN- approach and BMF-based approaches by Precision and Recall

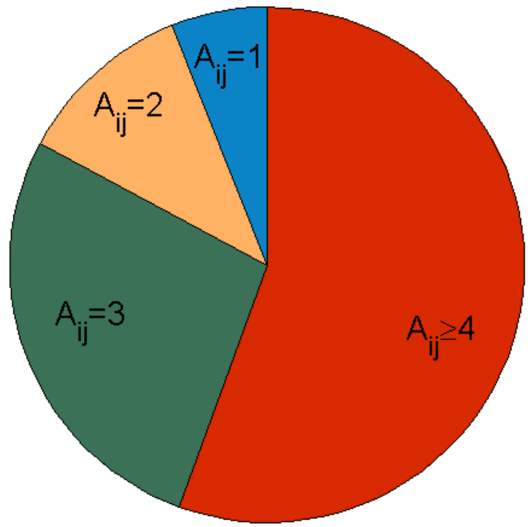




Experiments

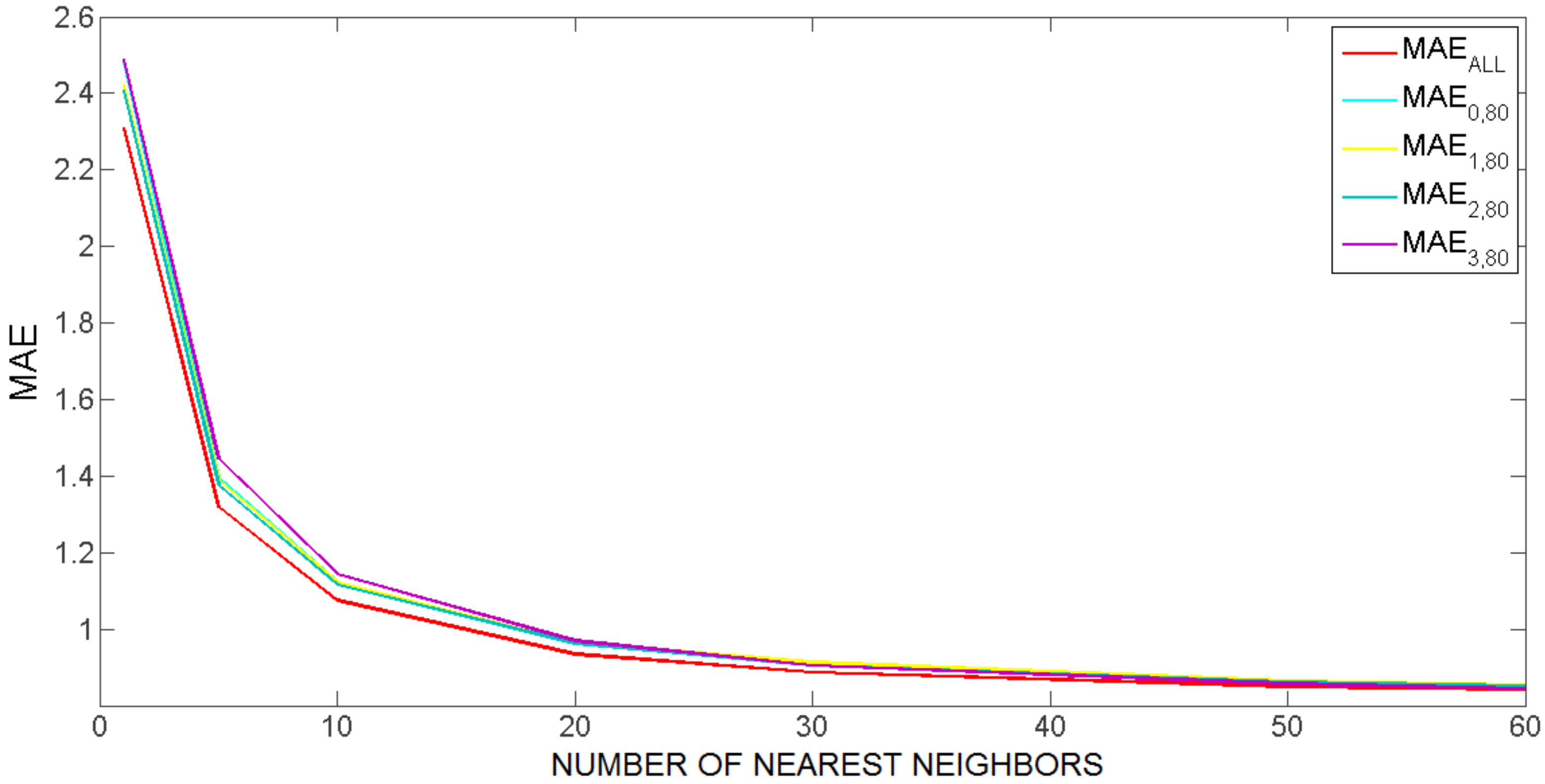
- Scaling influence on the recommendations quality for BMF in terms of MAE

1. $I_{ij} = 1$ if $R_{ij} > 0$, else $I_{ij} = 0$ (user i rates item j).
2. $I_{ij} = 1$ if $R_{ij} > 1$, else $I_{ij} = 0$.
3. $I_{ij} = 1$ if $R_{ij} > 2$, else $I_{ij} = 0$.
4. $I_{ij} = 1$ if $R_{ij} > 3$, else $I_{ij} = 0$.



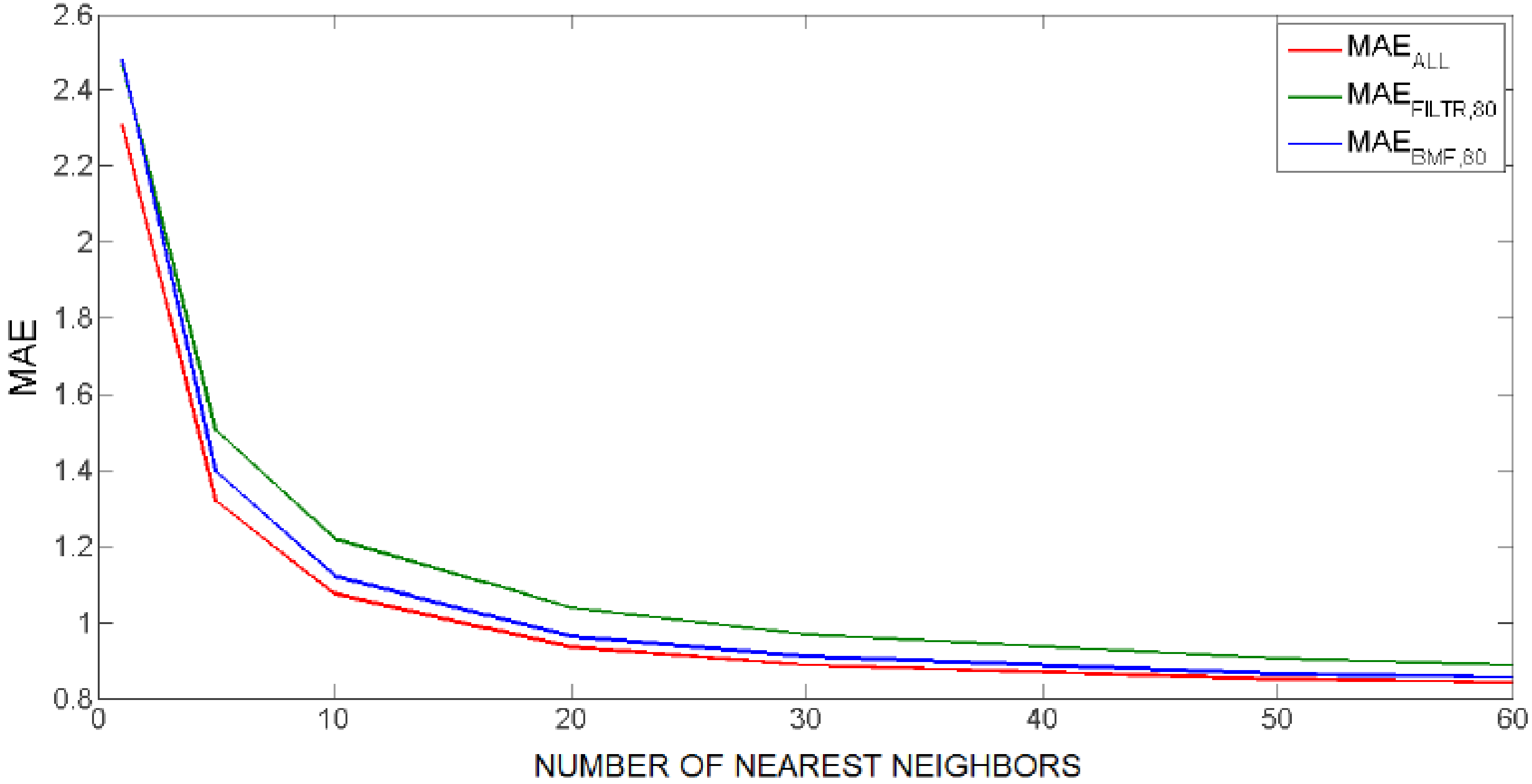
Experiments

- MAE dependence on scaling and number of nearest neighbors for 80% coverage.



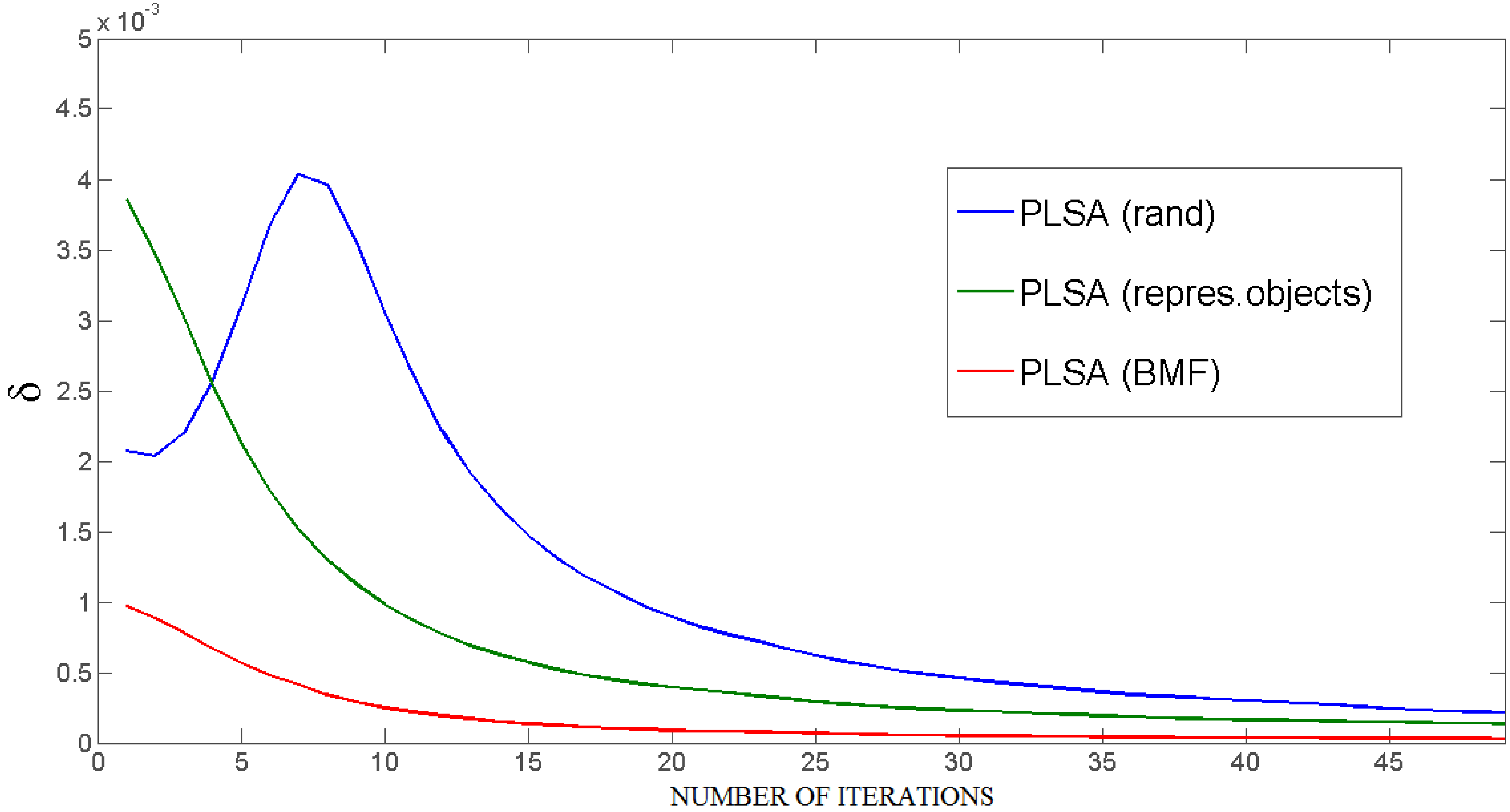
Experiments

- MAE dependence on data filtration algorithm and the number of nearest neighbors.



Experiments

- Speed up of PLSA convergence



Conclusion

- BMF-based RA is similar to state-of-the-art techniques in terms of MAE and demonstrates good Precision and Recall
- Probably low scalability is the main drawback of the approach
- BMF: $O(k|G||M|^3)$ versus SVD: $O(|G||M|^2+|M|^3)$

Future Prospects

- BMF-based RS in Triadic Case (e.g., folksonomy data)
- BMF-based RS for Graded and Ordinal Data
- BMF-based RS for simultaneous factorisation of user-features, user-items, and items-features matrices
- BMF and Least Square based imputation techniques
- Scalability Issues