CLUSTERED MONOTONE TRANSFORMS FOR RATING FACTORIZATION

Gaurush Hiranandani*, UIUC

Raghav Somani*, Microsoft Research, India

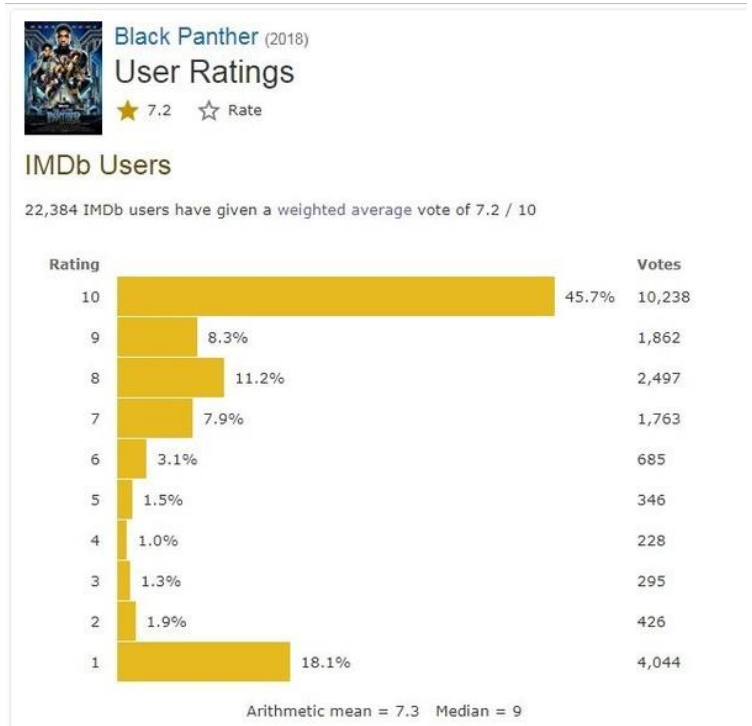
Oluwasanmi Koyejo, UIUC

Sreangsu Acharyya, Microsoft Research, India

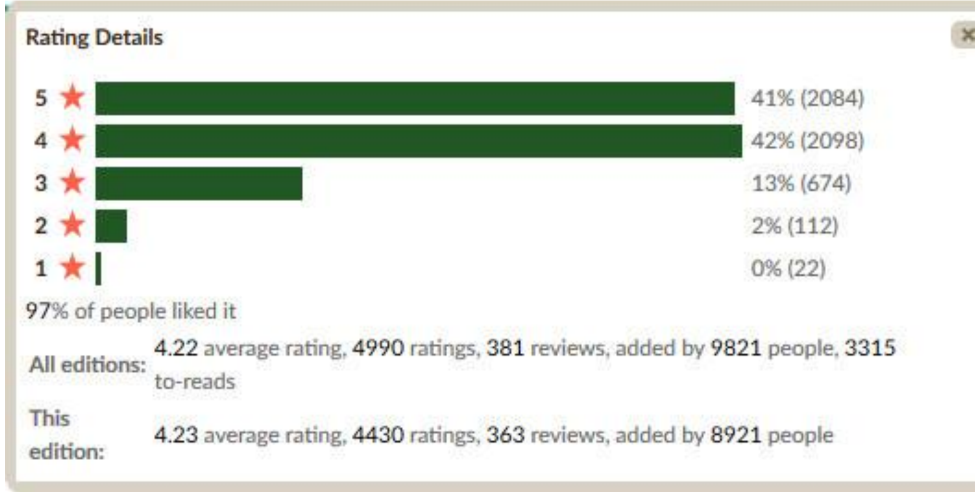
*Equal Contribution

MOTIVATION

IMDB

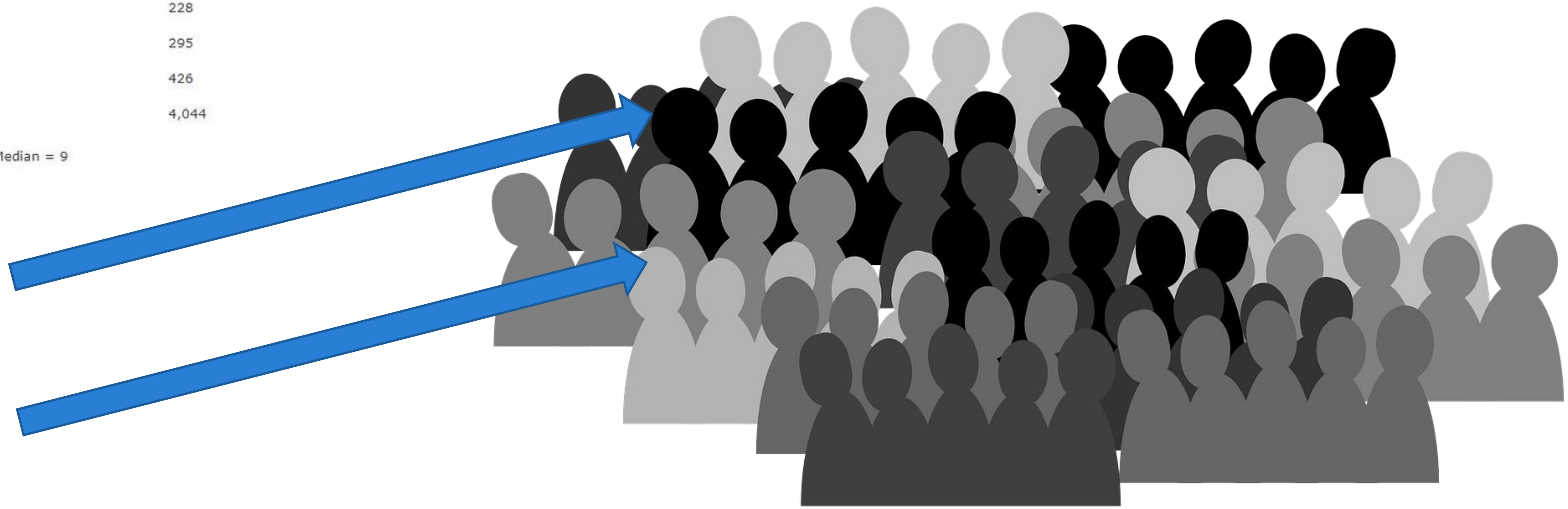


Goodreads

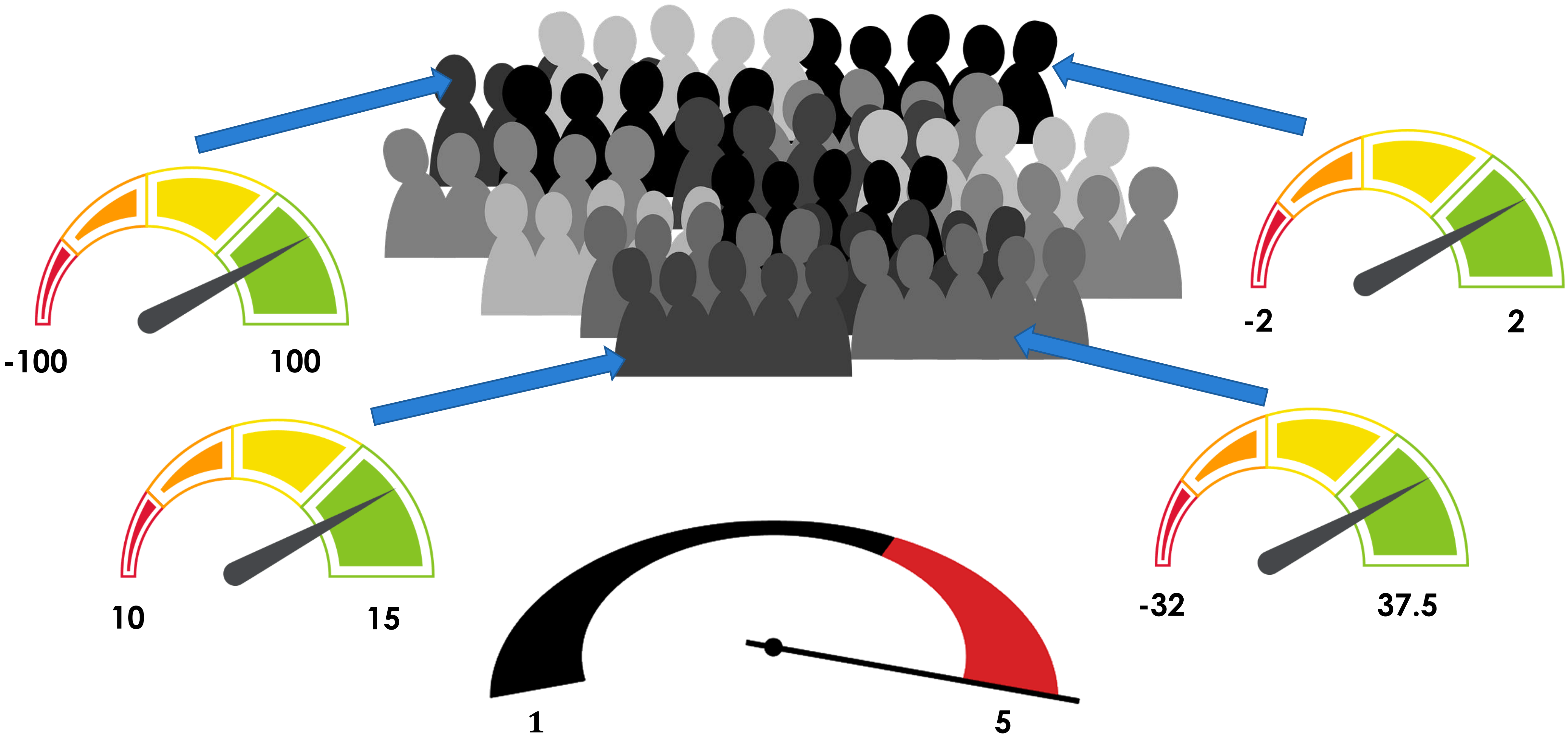


Extremists

Moderates



MOTIVATION



INTRODUCTION

$$n\left\{\overbrace{\begin{bmatrix} \quad & R & \quad \end{bmatrix}}^{m}\right. \approx n\left\{\overbrace{\begin{bmatrix} U \end{bmatrix}}^{d}\right. \left.\overbrace{\begin{bmatrix} \quad & V^T & \quad \end{bmatrix}}^{m}\right\} d$$

$$n\left\{\overbrace{\begin{bmatrix} \quad & \sigma^{-1}(R) & \quad \end{bmatrix}}^{m}\right. \approx n\left\{\overbrace{\begin{bmatrix} U \end{bmatrix}}^{d}\right. \left.\overbrace{\begin{bmatrix} \quad & V^T & \quad \end{bmatrix}}^{m}\right\} d$$

1-CMTRF

$$\min_{\substack{U,V,\\ \mathbf{r}\in\mathcal{R}_{L,\epsilon}}} \sum_i C_\phi(E_i\mathbf{r}\|V_i\boldsymbol{u}_i) + \frac{\lambda_u}{2}\sum_i\|\boldsymbol{u}_i\|^2 + \frac{\lambda_v}{2}\sum_j\|\boldsymbol{v}_j\|^2$$

Algorithm

$$\boldsymbol{x}^{t+1} = \arg\min_{\boldsymbol{x}\in\mathcal{R}_{L,\epsilon}} \sum_i C_\phi(E_i\boldsymbol{x}\|V_i^t\boldsymbol{u}_i^t)$$

(PAVA)

$$U^{t+1}, V^{t+1} = \arg\min_{U,V} \sum_i C_\phi(E_i\boldsymbol{x}^{t+1}\|V_i\mathbf{u}_i) + \frac{\lambda_u}{2}\sum_i||\boldsymbol{u}_i||^2 + \frac{\lambda_v}{2}\sum_j||\boldsymbol{v}_j||^2$$

(LS Solution)

N-CMTRF

$$\min_{\substack{U,V,\\ \{\mathbf{r}_i \in \mathcal{R}_{L,\epsilon}\}_{i=1}^N}} \sum_i C_\phi(E_i\mathbf{r}_i \| V_i \boldsymbol{u}_i) + \frac{\lambda_u}{2}\sum_i \|\boldsymbol{u}_i\|^2 + \frac{\lambda_v}{2}\sum_j \|\boldsymbol{v}_j\|^2$$

Algorithm

$$\boldsymbol{x}_i^{t+1} = \operatorname*{arg\,min}_{\boldsymbol{x}_i \in \mathcal{R}_{L,\epsilon}} \sum_i C_\phi(E_i \boldsymbol{x}_i \| V_i^t \boldsymbol{u}_i^t) \quad \forall i \text{ in parallel}$$

(PAVA)

$$U^{t+1}, V^{t+1} = \operatorname*{arg\,min}_{U,V} \sum_i C_\phi(E_i \boldsymbol{x}_i^{t+1} \| V_i \mathbf{u}_i) + \frac{\lambda_u}{2}\sum_i ||\boldsymbol{u}_i||^2 + \frac{\lambda_v}{2}\sum_j ||\boldsymbol{v}_j||^2$$

(LS Solution)

K-CMTRF

$$\min_{\substack{U,V,\{\mathbf{z}_i\}_{i=1}^{N}\\ \{\mathbf{r}_k\in\mathcal{R}_{L,\epsilon}\}_{k=1}^{K}}} \sum_{k=1}^{K}\sum_{i=1}^{N} z_{ik} C_\phi\left(E_i\mathbf{r}_k \| V_i\boldsymbol{u}_i\right) + \frac{\lambda_u}{2}\sum_i \|\boldsymbol{u}_i\|^2 + \frac{\lambda_v}{2}\sum_j \|\boldsymbol{v}_j\|^2$$

$$\mathbf{z}_i^{t+1} = \arg\min_{k\in[K]} C_\phi\left(E_i\mathbf{x}_k^t \| V_i^t\boldsymbol{u}_i^t\right) \;\; \forall\, i$$

(K-means style)

$$\boldsymbol{x}_k^{t+1} = \arg\min_{\boldsymbol{x}_k\in\mathcal{R}_{L,\epsilon}} \sum_{i:\mathbf{z}_{ik}^{t+1}=1} C_\phi\left(E_i\boldsymbol{x}_k \| V_i^t\boldsymbol{u}_i^t\right) \;\; \forall k \text{ in parallel}$$

(PAVA)

$$U^{t+1}, V^{t+1} = \arg\min_{U,V} \sum_{k=1}^{K}\sum_{i:\mathbf{z}_{ik}^{t+1}=1} C_\phi\left(E_i\boldsymbol{x}_k{}^{t+1} \| V_i\mathbf{u}_i\right) + \frac{\lambda_u}{2}\sum_i ||\boldsymbol{u}_i||^2 + \frac{\lambda_v}{2}\sum_j ||\boldsymbol{v}_j||^2$$

(LS Solution)

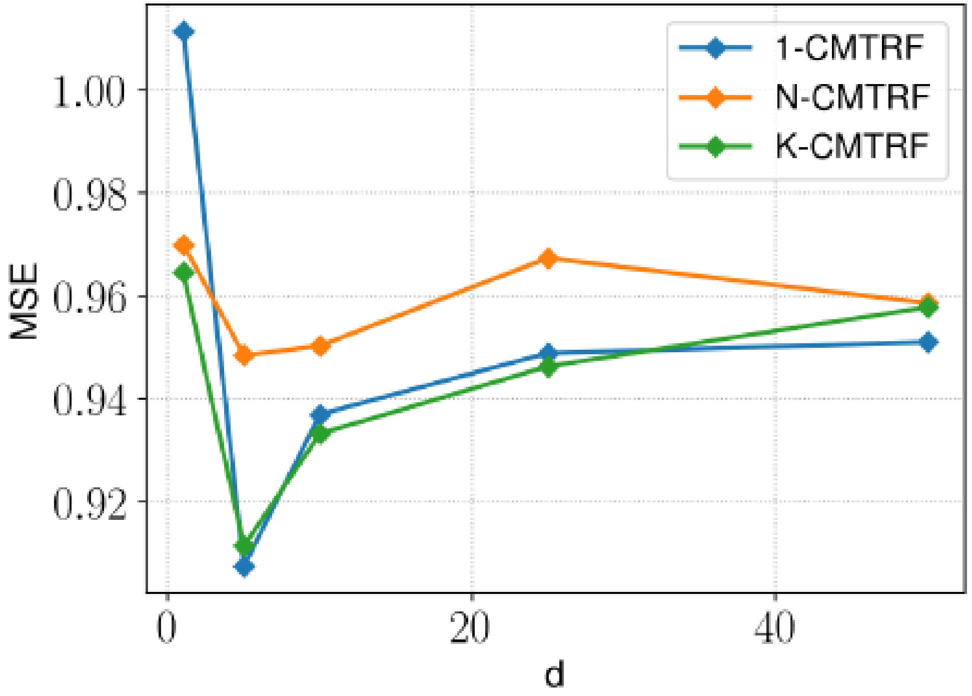
THEORETICAL RESULTS

- Alternating Minimization possible because objective functions are tri-convex
- Global solution guarantees for 1-CMTRF and N-CMTRF for square loss (cost) function
- Local solution guarantees for K-CMTRF for square loss (cost) function

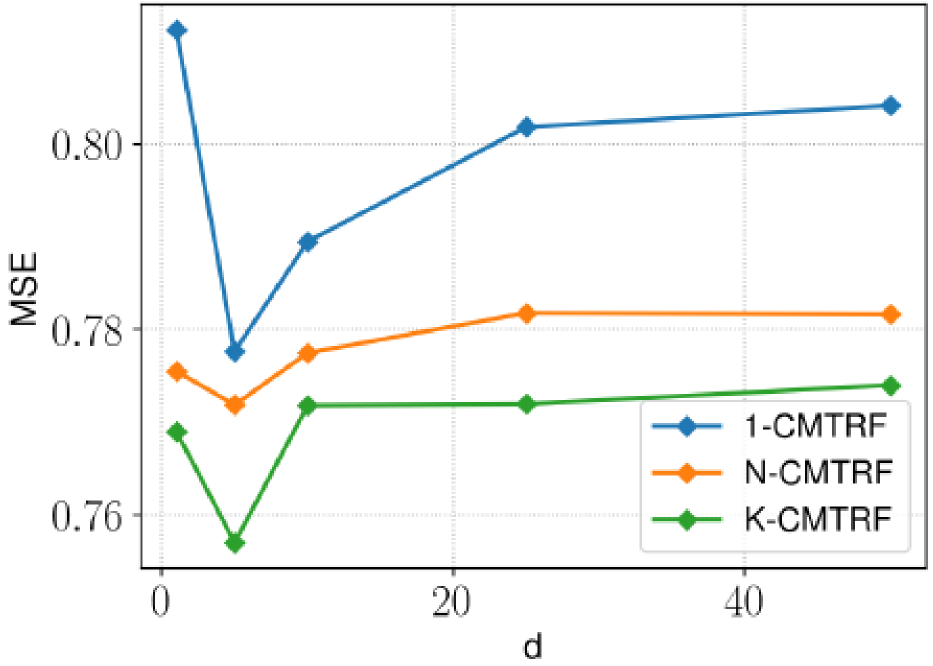
EMPIRICAL RESULTS

Chronological Split								
Metric	Data	MF	LMaFit	MMC	NNMF	1-C	N-C	K-C
MSE	ML100k	0.989	2.269	1.058	0.947	**0.908**	0.948	0.911
	ML1M	0.809	1.918	0.989	0.797	0.799	0.790	**0.778**
	ML10M	0.834	1.947	0.975	0.832	0.806	0.807	**0.799**
	GB	0.811	3.232	0.888	0.768	0.778	0.772	**0.757**
	Epinions	1.329	11.00	-	**1.113**	1.232	1.184	1.167
MAE	ML100k	0.989	1.173	0.830	0.755	0.759	0.758	**0.752**
	ML1M	0.708	1.058	0.777	0.700	0.723	**0.693**	0.778
	ML10M	0.712	0.988	0.791	0.698	0.708	0.697	**0.691**
	GB	0.701	1.418	0.753	0.691	0.694	0.677	**0.670**
	Epinions	0.864	2.886	-	**0.800**	0.845	0.812	0.811
Uniform Split								
Metric	Data	MF	LMaFit	MMC	NNMF	1-C	N-C	K-C
MSE	Douban	0.543	1.462	0.846	0.542	0.541	0.534	**0.534**
	Flixster	0.870	6.789	1.52	0.765	0.869	0.757	**0.729**
	ML100k_u	0.890	2.240	1.072	0.876	0.846	0.833	**0.833**
MAE	Douban	0.575	0.934	0.720	0.576	0.575	0.570	**0.570**
	Flixster	0.704	2.005	1.015	0.667	0.710	**0.623**	0.630
	ML100k_u	0.749	1.160	0.808	0.744	0.728	**0.720**	0.721

EMPIRICAL RESULTS



Movielens



Epinions

CONCLUSIONS

- Proposed Clustered Monotonic Transformation for Rating Factorization: regression upto unknown monotonic transform over unknown population segments
- Three algorithms
- Unique solution for squared loss and local solution for general Bregman Divergences