Discrete Content-aware Matrix Factorization

Rui Liu, University of Electronic Science and Technology of China

Joint work with Defu Lian (UESTC), Yong Ge (UA), Kai Zheng (UESTC), Xing Xie (Microsoft Research) and Longbing Cao (UTS)



Outline

Motivation

- Proposed framework
- Experiment
- Conclusion

Motivation

Google News Recommendation

Tasks: recommends new articles based on click and search History

Scale:

- millions of users
- millions of articles •

News

Top Stories London

Donald Trump

Grenfell Tower

James Comey

Steve Scalise

Robert Mueller

Suggested for you

World

U.S.

Business

Technology

Sports Science

Health

Spotlight

Entertainment

Amazon.com, Inc.

Whole Foods Market

Sydney, New South Wa.

U.S. Open

Jefferson Sessions

Top Stories



U.S. edition -

Bill Cosby Mistrial: Hollywood Reporter Leg

Hollywood Reporter - 7 hours ago 🛛 😚 📂 📝 🖪

"Are either of you surprised? Can't say I am," says THR senior editor I sense: "It's always nearly impossible to predict

Camille Cosby and Others React to the Bill Cosby Jury Verdict New Yo Lena Dunham Sends Message of Support to Survivors of Sexual Assa

Highly Cited: Bill Cosby trial: Mistrial declared after jury deadlocks CN Opinion: Bill Cosby was largely irrelevant in black America even before Wikipedia: Bill Cosby sexual assault allegations

See realtime coverage







Number of missing US sailors found dead after collision with me Washington Post - 1 hour ago TOKYO - A number of the seven U.S. Navy sailors missing after a collision betwee



Trump's Cuba policy tries to redefine 'good' US tourism. That in

The American traveler in Cuba - sweating, disoriented and probably a bit woozy fror



Trump breaks weekend streak with visit to Camp David

The Hill - 2 hours ago President Trump broke from a self-established tradition on Saturday, foregoing a trip

The Latest: UK's May meets fire survivors, faces criticism Washington Post - 6 hours ago



LONDON - The Latest on the London high-rise fire (all times local):. 8:40 p.m.. Briti 15 survivors and community leaders at ...



Scalise Shooting: GOP congressman upgraded from 'critical' to Fox News - 2 hours ago

The Republican congressman who was wounded in a shooting on Wednesday saw hospital officials.

Suggested for you »

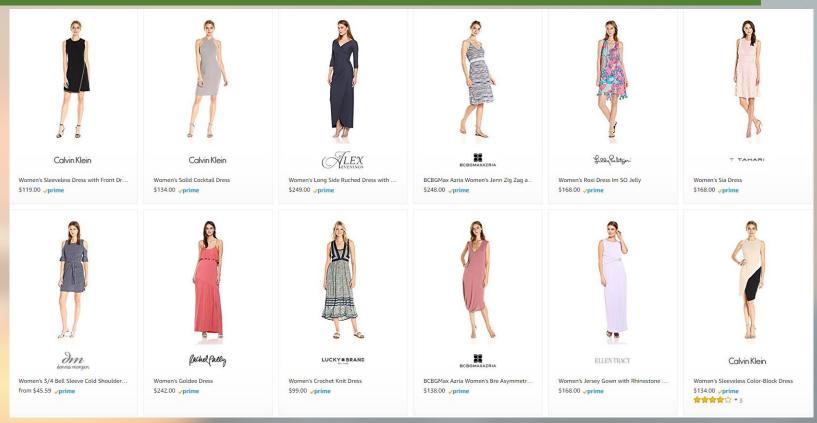


Warriors' Steve Kerr on Steph Curry: 'I'm such an idiot' The Mercury News - 7 hours ago





Motivation



Amazon Product Recommendation

Tasks: recommends new articles based on click and search History

Scale:

- 300 million users
- 480 million products

Motivation

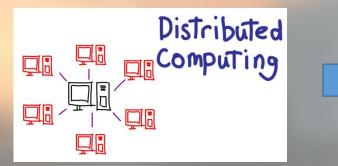
- How to generate immediate recommendation?
 - Pre-computing top-k preferred items for users?



User interests are evolving

New items are emerging

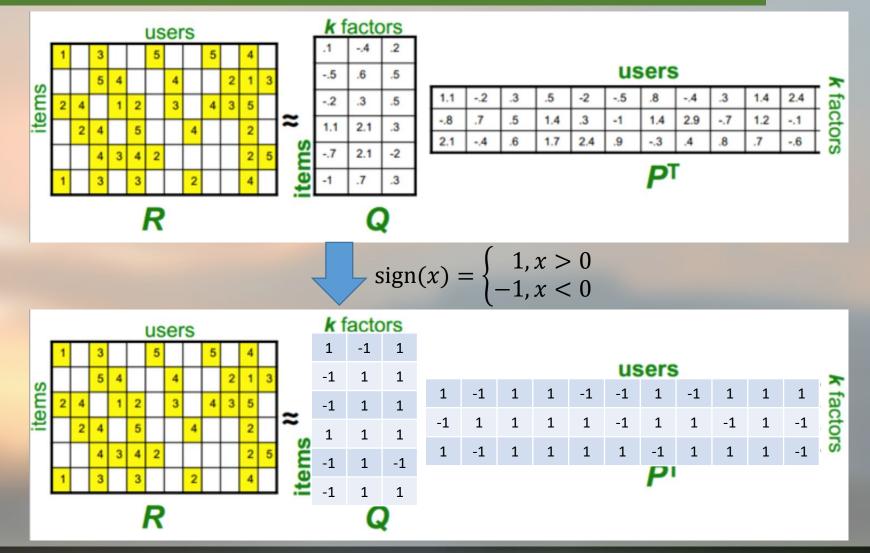
Distributed/parallel computing





A lot of machines

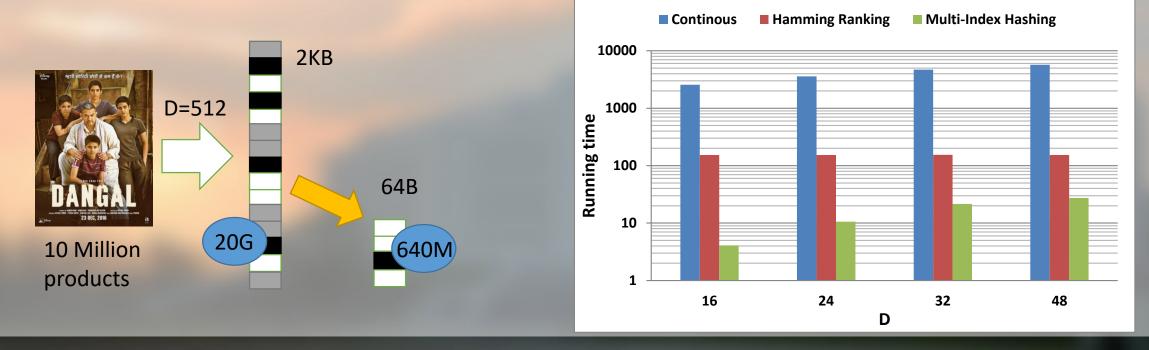
Economic and Effective Way



Economic and Effective Way

$$\langle p_u, q_i \rangle = \sum_{d=1}^{D} p_{ud} q_{id}$$
 $p_u, q_i \in \mathbb{R}^{D}$

 $\langle \boldsymbol{\phi}_{u}, \boldsymbol{\psi}_{i} \rangle = 2H(\boldsymbol{\phi}_{u}, \boldsymbol{\psi}_{i}) - D \quad \boldsymbol{\phi}_{u}, \boldsymbol{\psi}_{i} \in \{\pm 1\}^{D}$



Related Work

- Dimension Reduction Techniques for Recommendation Systems
 - Matrix Factorization (KDD'11)
 - Map users and items into low-dimension latent space
 - Content-aware Matrix Factorization (RecSys'13, ICDM'15)
 - Take content information into account, help solve cold-start problem
 - Weighted Regularized Matrix Factorization
 - Work well in collaborative filtering for implicit feedback
- Distributed Computing Techniques for RS
 - Large-scale Parallel Collaborative Filtering
 - Parallel update to improve efficiency

Related Work

- Learning Hashing Codes for Recommendation Systems
 - Binary Code learning method for Collaborative Filtering (KDD'12)
 - Obtained binary codes preserve preference of users
 - Preference Preserving Hashing (SIGIR'14)
 - Come up with a novel quantization algorithm
 - Discrete Collaborative Filtering (SIGIR'16)
 - Tackle discrete optimization directly and efficiently
 - 1. Cold-start problem
 - 2. Implicit feedback
 - 3. Classification

Task

• DCMF: a joint framework

- Incorporate content information into the total framework
 - Solve cold-start problem
- Combine the Weighted Regularized Matrix Factorization
 - Solve for implicit feedback
- Find a solution for logit loss
 - Solve for classification
- A direct discrete optimization model
 - Update efficiently

Outline

- Motivation
- Proposed framework
- Experiment
- Conclusion

Total Framework

$$\min_{\substack{\Phi,\Psi,P,Q\\ U,V}} \sum_{i,j\in\Omega} \ell(r_{ij},\phi'_{i}\psi_{j}) + \alpha_{1} \|\Phi - P\|_{F}^{2} + \alpha_{2} \|\Psi - Q\|_{F}^{2}$$
 balance and decorrelation constraints balance and decorrelation balance and decorrelation constraints balance and decorrelation balance and decorrelation constraints balance and decorrelation c

Alternative Update Rule

$$\phi_{id}^{*} = sgn\left(K(\hat{\phi}_{id}, \phi_{id})\right)_{K(x, y)} = \begin{cases} x, x \neq 0 \\ y, x = 0 \end{cases}$$
$$\hat{\phi}_{id} = \sum_{j \in \mathbb{I}_{i}} (r_{ij} - \hat{r}_{ij} + \phi_{id}\psi_{jd})\psi_{jd} + \alpha_{1}p_{id}$$
$$(r_{ij}/4 - \lambda(\hat{r}_{ij})\hat{r}_{ij} + \lambda(\hat{r}_{ij})\phi_{id}\psi_{jd})\psi_{jd}$$
$$+\lambda_{1}u'_{d}x_{i} - \beta\phi'_{i}\Psi'\psi_{d} + \beta N\phi_{id}$$

$$\mathbf{P} = \sqrt{M} \begin{bmatrix} S_P, \widehat{S}_P \end{bmatrix} \begin{bmatrix} T_P, \widehat{T}_P \end{bmatrix}'$$
$$S_P \in \mathbb{R}^{M \times \widetilde{D}} \quad \widehat{S}_P \in \mathbb{R}^{M \times (D - \widetilde{D})}$$
$$T_P \in \mathbb{R}^{D \times \widetilde{D}} \quad \widehat{T}_P \in \mathbb{R}^{D \times (D - \widetilde{D})}$$
$$\mathbf{U} = \left(\mathbf{X}' \mathbf{X} + \frac{\gamma_1}{\lambda_1} \mathbf{I}_F \right)^{-1} \mathbf{X}' \mathbf{\Phi}$$

$$\psi_{jd}^* = sgn\left(K(\hat{\psi}_{jd},\psi_{jd})\right)$$

$$\hat{\psi}_{jd} = \sum_{i \in \mathbb{I}_j} \frac{(r_{ij} - \hat{r}_{ij} + \phi_{id}\psi_{jd})\phi_{id} + \alpha_2 q_{jd}}{(r_{ij}/4 - \lambda(\hat{r}_{ij})\hat{r}_{ij} + \lambda(\hat{r}_{ij})\phi_{id}\psi_{jd})\phi_{id}} + \lambda_2 v'_d y_j - \beta \psi'_j \Phi' \phi_d + \beta M \psi_{jd}}$$

Υ'Ψ

$$\mathbf{Q} = \sqrt{N} \begin{bmatrix} S_Q, \widehat{S}_Q \end{bmatrix} \begin{bmatrix} T_Q, \widehat{T}_Q \end{bmatrix}'$$

$$S_Q \in \mathbb{R}^{N \times \widetilde{D}} \qquad \widehat{S}_Q \in \mathbb{R}^{N \times (D - \widetilde{D})}$$

$$T_Q \in \mathbb{R}^{D \times \widetilde{D}} \qquad \widehat{T}_Q \in \mathbb{R}^{D \times (D - \widetilde{D})}$$

$$(\qquad) -1$$

 $\mathbf{V} = \left(\mathbf{Y}'\mathbf{Y} + \frac{\gamma_2}{\lambda_2}\mathbf{I}_L\right)$

Outline

- Motivation
- Proposed framework
- Experiment
- Conclusion

Dataset and Metric

- ✓ MovieLens, classic MovieLens 10M dataset
- ✓ Yelp, the latest Yelp Challenge dataset
- Amazon, a subset of product reviews and metadata for Amazon books

Dataset	#users	#items	#ratings	Density
MovieLens	69,838	8,940	9,983,758	1.60%
Yelp	13,679	12,922	640,143	0.36%
Amazon	35,151	33,195	1,732,060	0.15%

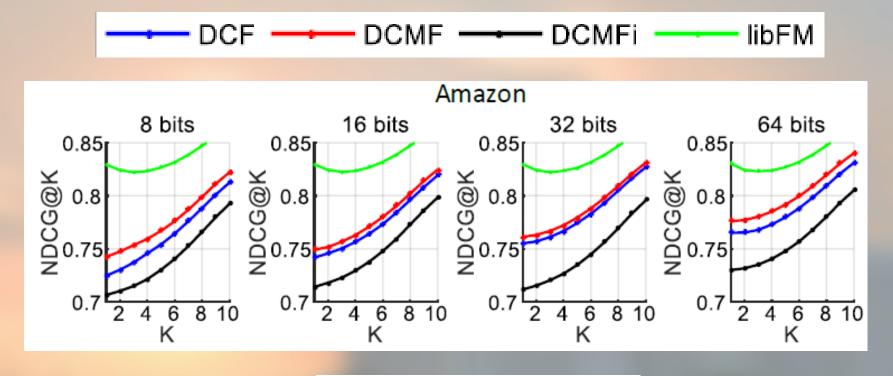
Data statistics

✓ Metric: NDCG for regression MPR for classification

Baselines

• DCF

- Hashing-based collaborative filtering.
- Outperforms almost all two-stage binary code learning methods for collaborative filtering including BCCF, PPH, CH.
- libFM:
 - Feature-based recommendation system.
 - Has achieved the best sole-model for the track I challenge-link prediction-of KDD-Cup 2012.
 - Supports both regression and classification tasks of recommendation.

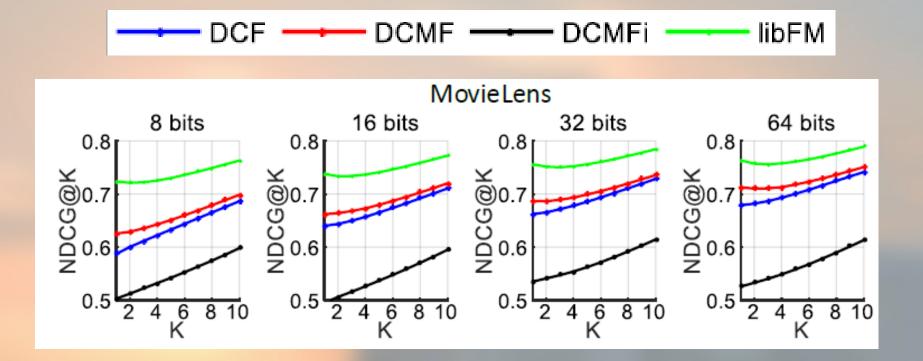


In-matrix regression

Add content information

 \rightarrow performance \uparrow

 Effective discrete optimization algorithm

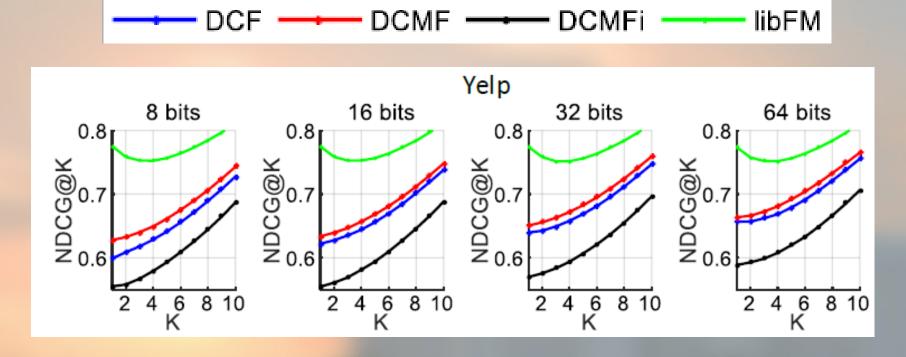


Add content information

 \rightarrow performance \uparrow

 Effective discrete optimization algorithm

In-matrix regression

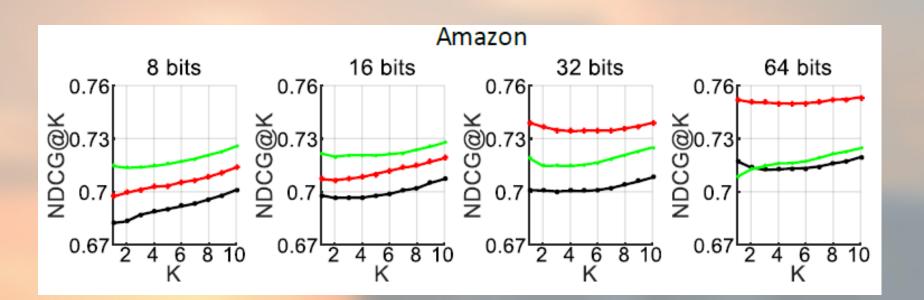


In-matrix regression

Add content information

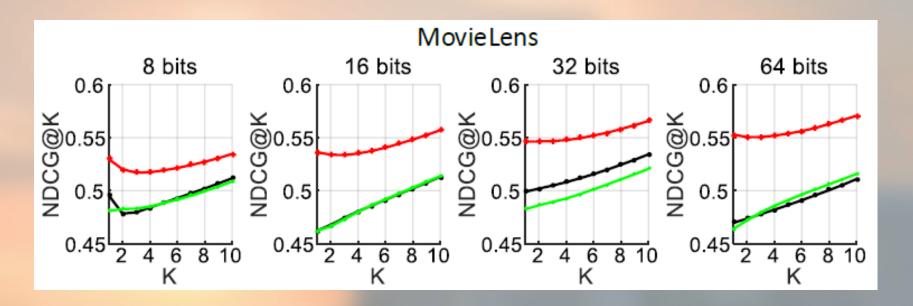
 \rightarrow performance \uparrow

• Effective discrete optimization algorithm



Out-matrix regression

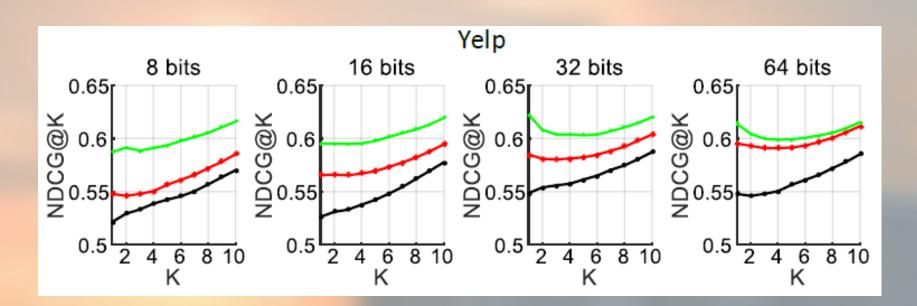
- Address cold-start problem well
- Effective discrete optimization algorithm



Address cold-start problem well

 Effective discrete optimization algorithm

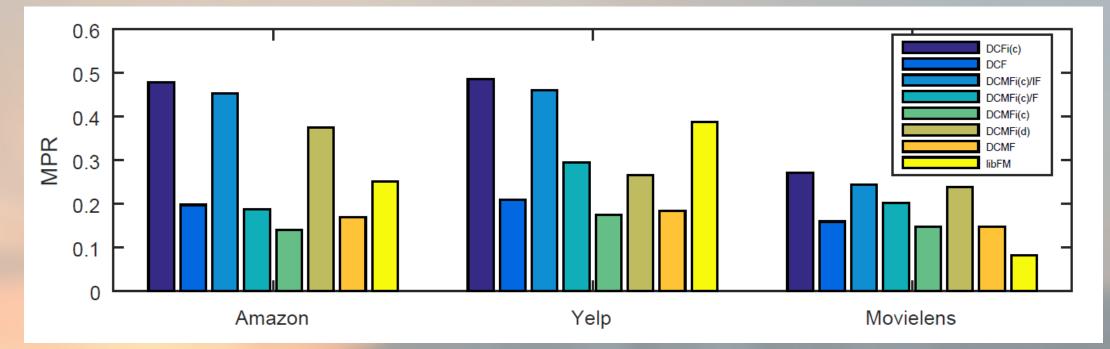
Out-matrix regression



- Address cold-start problem well
- Effective discrete optimization algorithm

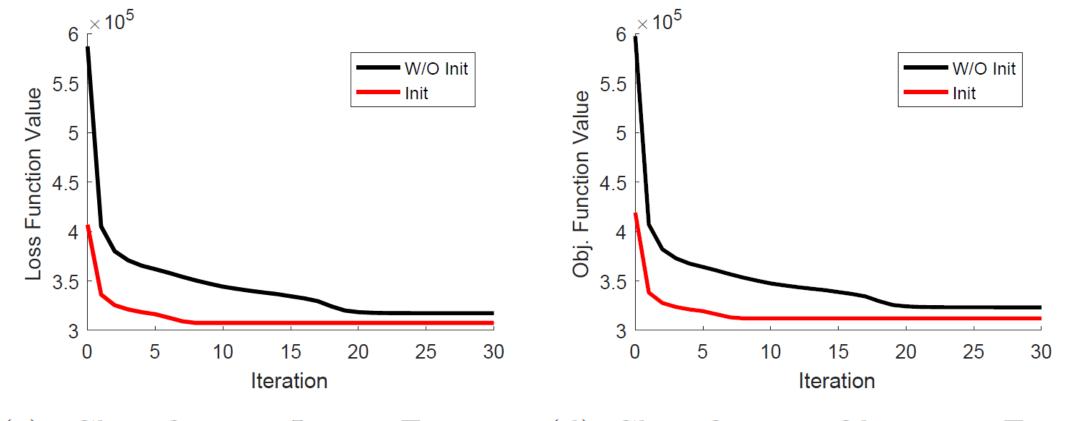
Out-matrix regression

Comparison with baselines - Classification



- Effectiveness of content information and iteration regularization
- ✓ Benefit of the use of logit loss
- ✓ Superiority of DCMF to DCF
- ✓ Validity of the proposed discrete optimization algorithm
- ✓ Superiority of DCMF to libFM on relatively sparse datasets

Convergence Study



(c) Classification-Loss Function (d) Classification-Objective Func-Value tion Value

Outline

- Motivation
- Proposed framework
- Experiment
- Conclusion

Conclusion

- Proposed a new framework called DCMF to hash users and items with content information in both regression and classification tasks.
- Developed an efficient discrete optimization algorithm for tackling discretized constraints as well as interaction regularization.
- Proved superiority in 3 public datasets

Thank you!

Rui Liu, UESTC ruiliu011@gmail.com