Discrete Content-aware Matrix Factorization

Defu Lian, Rui Liu, Yong Ge, Kai Zheng, Xing Xie, Longbing Cao

Big Data Research Center and School of Computer Science and Engineering, University of Electronic Science and Technology of China;
Eller School of Management, University of Arizona; Microsoft Research; University of Technology Sydney

KDD2017

ABSTRACT

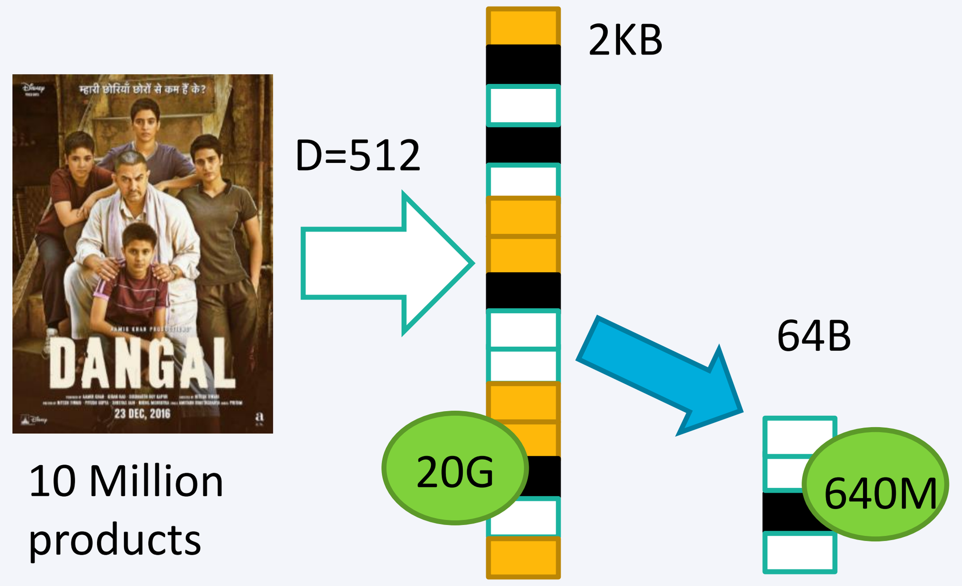
Precisely recommending relevant items from massive candidates to a large number of users is an indispensable yet computationally expensive task in many online platforms (e.g., Amazon.com and Netix.com). A promising way is to project users and items into a Hamming space and then recommend items via Hamming distance. However, previous studies didn't address the cold-start challenges and couldn't make the best use of preference data like implicit feedback. To fill this gap, we propose a Discrete Content-aware Matrix Factorization (DCMF) model, 1) to derive compact yet informative binary codes at the presence of user/item content information; 2) to support the classification task based on a local upper bound of logit loss; 3) to introduce an interaction regularization for dealing with the sparsity issue. We further develop an efficient discrete optimization algorithm for parameter learning. Based on extensive experiments on three real-world datasets, we show that DCMF outperforms the state-of-the-arts on both regression and classification tasks.

INTRODUCTION

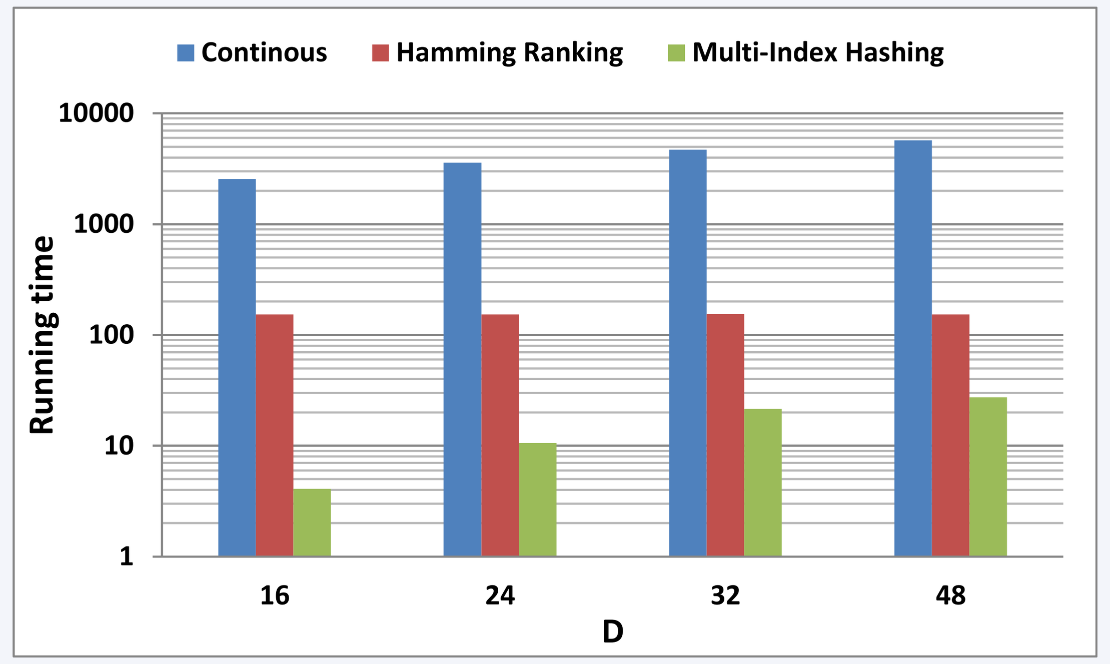
Recommender systems aim to recommend relevant items (e.g., products and news) to users via mining and understanding their preferences for items. Thanks to the development of recommendation techniques over the past decades, they have been widely used in various web services such as Amazon.com and eBay.com for improving sale of products and increasing click-through rate of advertisements. However, within most of these web services, the number of customers and products is dramatically growing, making recommendation more challenging than ever before. Consequently, it is challenging to generate immediate response to fond out potentially-preferred products for customers via analyzing large-scale yet sparse browsing, purchasing and searching data.

Why hashing?

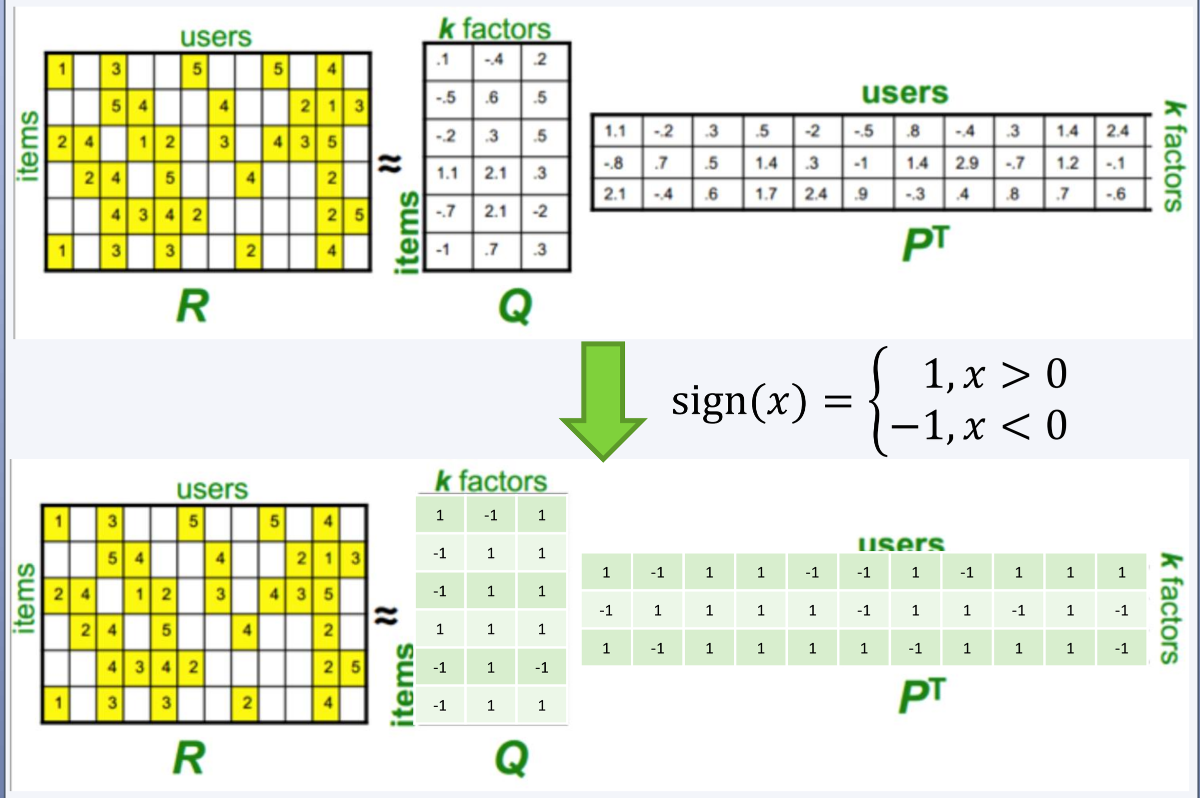
1. Save memory consumption



2. Reduce running time



Since matrix factorization often generates the best sole-model performance, we choose it for this study. Based on the learned latent factors of both users and items, we apply the sign function to map them into one or minus one.



$$\mathrm{sign}(x) = \begin{cases} 1, x > 0 \\ -1, x < 0 \end{cases}$$

CONTRIBUTIONS

- We study how to hash users and items at the presence of their respective content information for fast recommendation in both regression and classification tasks.
- We develop an efficient discrete optimization algorithm for tackling discretization, balanced and de-correlated constraints as well as interaction regularization.
- Through extensive experiments on three public datasets, we show the superiority of the proposed algorithm to the state-of-the-arts.

FRAMEWORK

$$\sum_{i,j\in\Omega} \ell(r_{ij}, \phi_i'\psi_j) + \alpha_1\|\mathbf{\Phi} - \mathbf{P}\|_F^2 + \alpha_2\|\mathbf{\Psi} - \mathbf{Q}\|_F^2 + \beta\sum_{i,j}(\phi_i'\psi_j)^2$$
$$+\lambda_1\|\mathbf{\Phi} - \mathbf{XU}\|_F^2 + \lambda_2\|\mathbf{\Psi} - \mathbf{YV}\|_F^2 + \gamma_1\|\mathbf{U}\|_F^2 + \gamma_2\|\mathbf{V}\|_F^2$$

$$s.t. \mathbf{\Phi} \in \{\pm 1\}^{M\times D}, \mathbf{\Psi} \in \{\pm 1\}^{N\times D}$$
$$\mathbf{1}'\mathbf{P} = 0, \mathbf{1}'\mathbf{Q} = 0, \mathbf{P}'\mathbf{P} = M\mathbf{I}, \mathbf{Q}'\mathbf{Q} = N\mathbf{I}$$

EXPERIMENTS & RESULTS

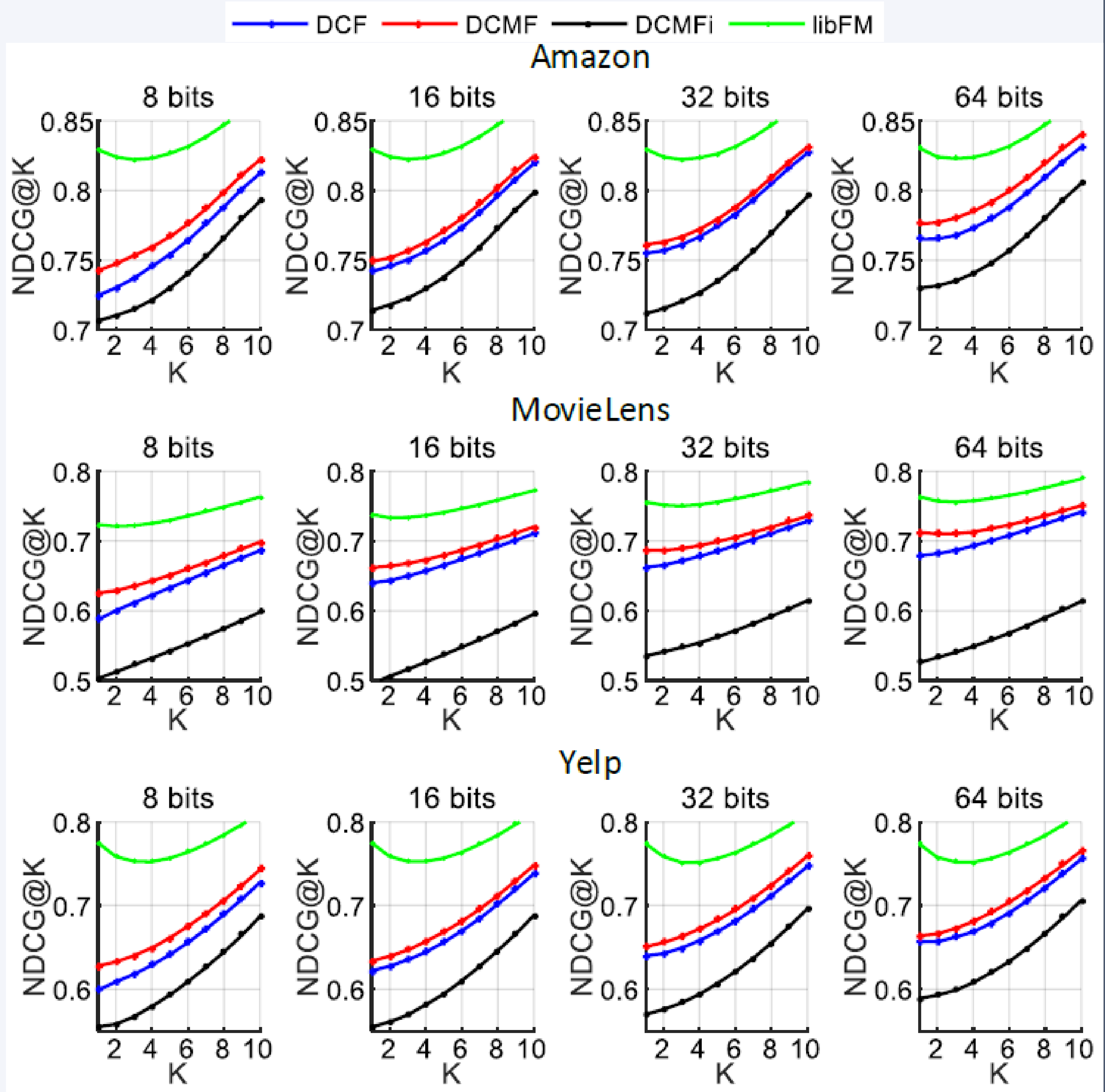
Statistics of Dataset

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%

Metric

- NDCG for regression
- MPR for classification

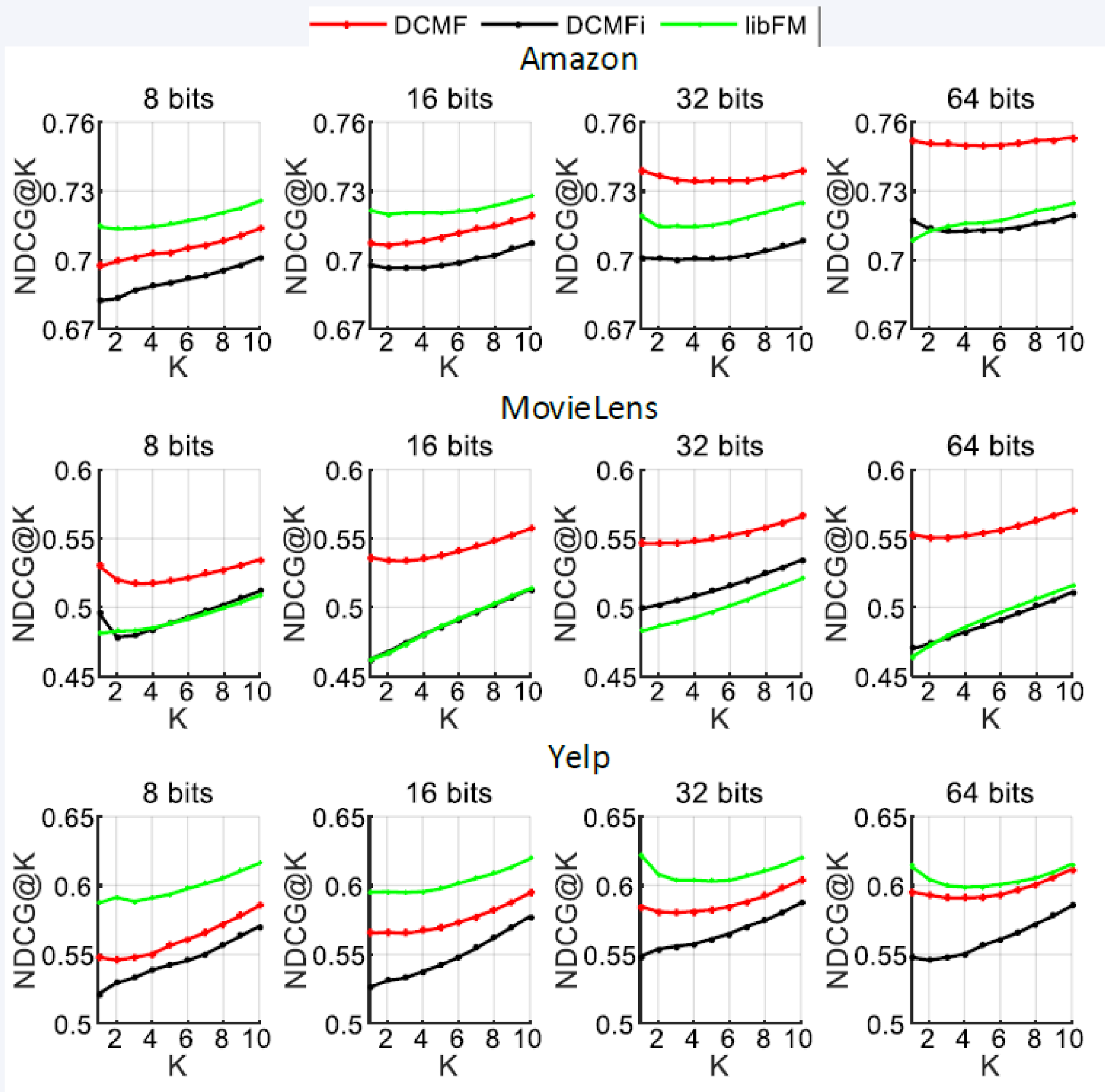
In-matrix Regression



✓ DCMF is superior to DCF

✓ Effective mixed-Integer optimization algorithm

Out-matrix Regression

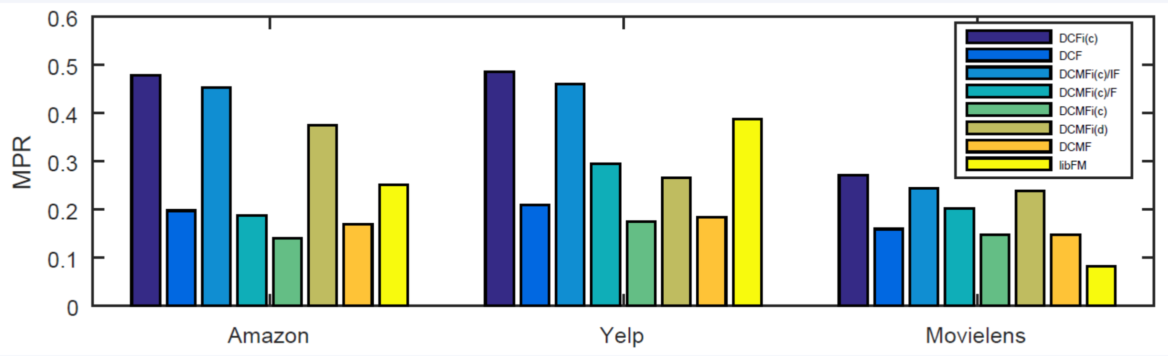


(b) Out-matrix regression

✓ Address cold-start problem well

✓ Effective discrete optimization algorithm

Classification



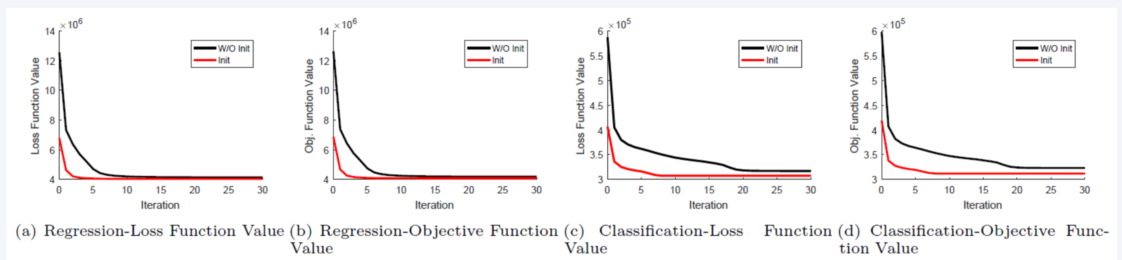
✓ Effectiveness of content information and interaction regularization

✓ Superiority of DCMF to DCF

✓ Effectiveness of mixed-integer optimization

✓ Superiority of DCMF to LibFM on sparse datasets

Convergence curve



CONCLUSION

In this paper, we propose Discrete Content-aware Matrix Factorization for investigating how to learn informative and compact hash codes for users and items at the presence of content information, and extend the recommendation task from regression to classification. Simultaneously, we suggest an interaction regularization, which penalizes non-zero predicted preference, for dealing with the sparsity challenge. We then develop an efficient discrete optimization algorithm for learning hash codes for users and items. The evaluation results of the proposed algorithm on three public datasets not only demonstrates the capability of the proposed algorithm for incorporating content information, but also outperforms the state-of-the-art hashing-based collaborative filtering algorithms on both regression and classification tasks. And it is interestingly observed that the interaction regularization could greatly improve the recommendation performance when the user-item matrix is sparse, verifying the effect of the interaction regularization at addressing the sparsity issue.

ACKNOWLEDGEMENT

The work is supported by the National Natural Science Foundation of China (61502077,61631005,61602234,61572032) and the Fundamental Research Funds for the Central Universities (ZYGX2014Z012,ZYGX2016J087).

SELECTED REFERENCES

Deepak Agarwal and Bee-Chung Chen. 2009. Regression-based latent factor models. In *Proceedings of KDD'09*. ACM, 19-28.

Rainer Gemulla, Erik Nijkamp, Peter J Haas, and Yannis Sismanis. 2011. Large-scale matrix factorization with distributed stochastic gradient descent. In *Proceedings of KDD'11*. ACM, 69-77.

Defu Lian, Cong Zhao, Xing Xie, Guangzhong Sun, Enhong Chen, and Yong Rui. 2014. GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of KDD'14*. ACM, 831-840.

Steffen Rendle. 2012. Factorization machines with libFM. *ACM Trans. Intell. Syst. Tech.* 3, 3 (2012), 57.

Hanwang Zhang, Fumin Shen, Wei Liu, Xiangnan He, Huanbo Luan, and Tat-Seng Chua. 2016. Discrete collaborative filtering. In *Proceedings of SIGIR'16*, Vol. 16.

Zhiwei Zhang, Qifan Wang, Lingyun Ruan, and Luo Si. 2014. Preference preserving hashing for effcient recommendation. In *Proceedings of SIGIR'14*. ACM, 183-192.

CONTACT

dove.ustc@gmail.com, ruiliu011@gmail.com