The Minority Matters: A Diversity-Promoting Collaborative Metric Learning Algorithm

Shilong Bao, Qianqian Xu, Zhiyong Yang, Yuan He,
Xiaochun Cao, Qingming Huang

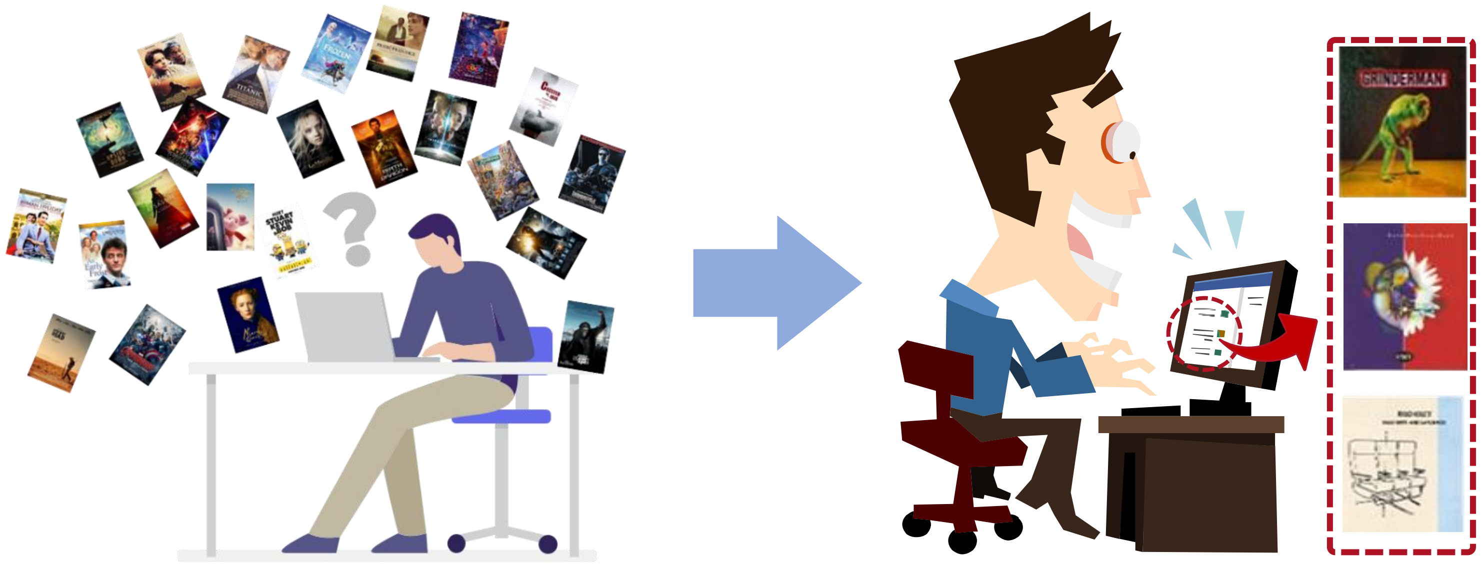
NeurIPS'22 (Oral)

Outlines

- Background
- Methodology
- Experiments
- Conclusion

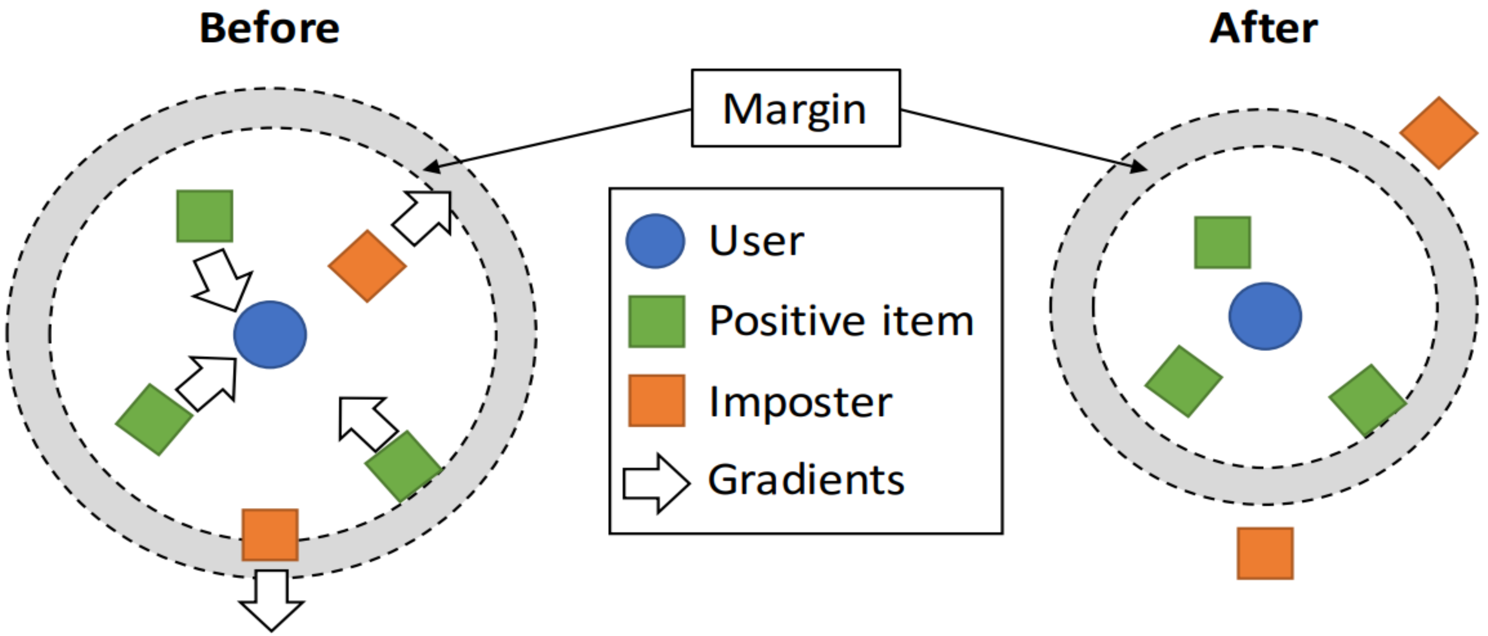
Background

- Recommendation Systems (RS) have become an essential building block in eCommerce to alleviate information overload.
- RS makes recommendation predictions based on users' historical data.



Background

- Collaborative Metric Learning (CML) has recently emerged as a popular method in RS:



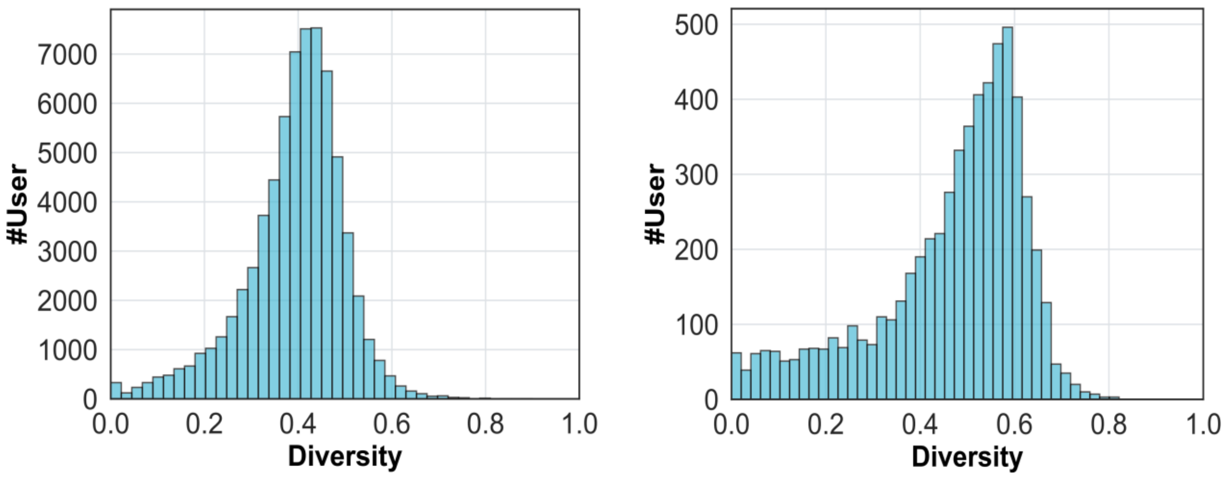
The illustration of Collaborative Metric Learning (CML) from [1].

[1] Hsieh, Cheng-Kang, et al. "Collaborative metric learning." WWW.

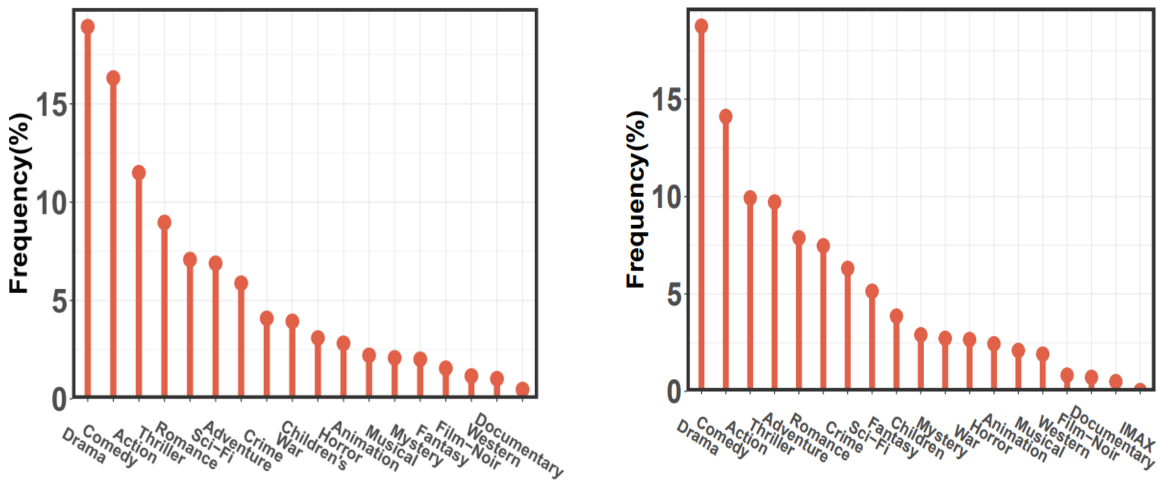
NEURAL INFORMATION PROCESSING SYSTEMS

Background

- Practical RS often has the following challenging properties:
  - ✓ Users have multiple interests and preferences
  - ✓ The item category distribution is imbalanced



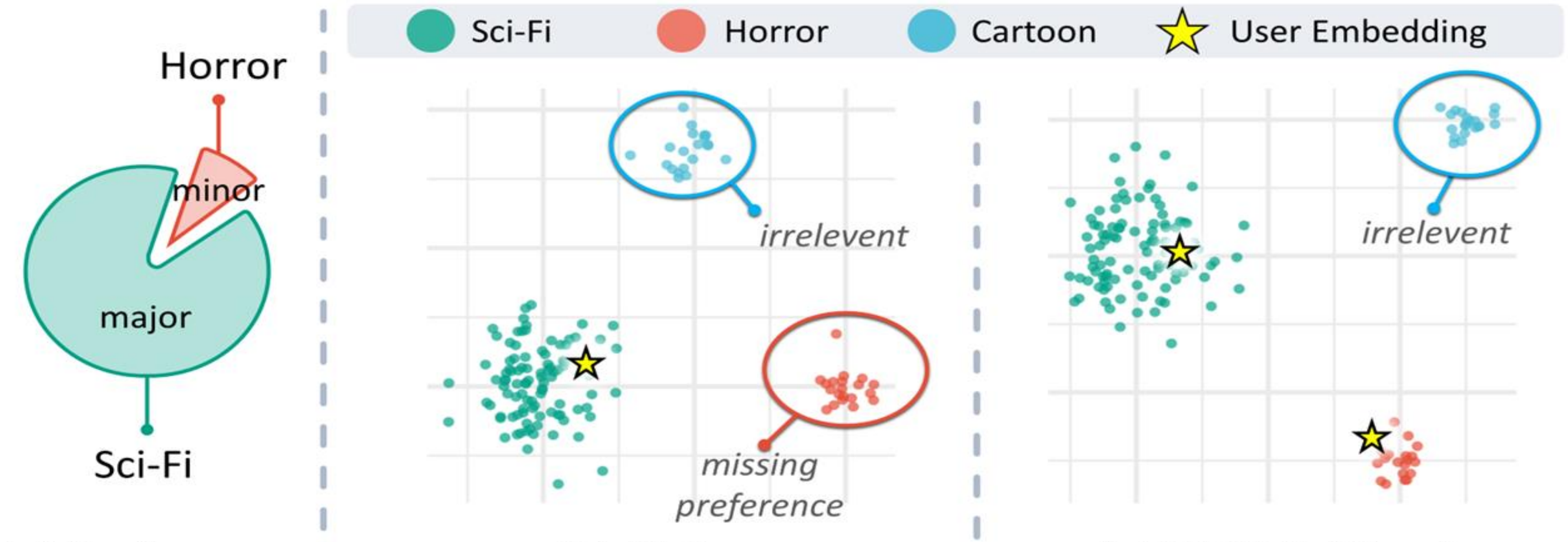
Statistics of preference diversity on MovieLens-1M/10M datasets.



The item category distribution on MovieLens-1M/10M datasets.

Background

- CML would pay more attention to the majority preference of the user and overlook the minority interest.



(a) Preference (b) CML (c) DPCML (Ours)

Question1: How to develop a Diversity-Promoting CML to consider all preferences of users?

Methodology

- DPCML assign a multiple set of vectors for each user
- User preference toward an item is defined as follows:

$$s(u_i, v_j) = \min_{c \in [C]} \|\boldsymbol{g}_{u_i}^c - \boldsymbol{g}_{v_j}\|^2, \forall v_j \in \mathcal{I}$$

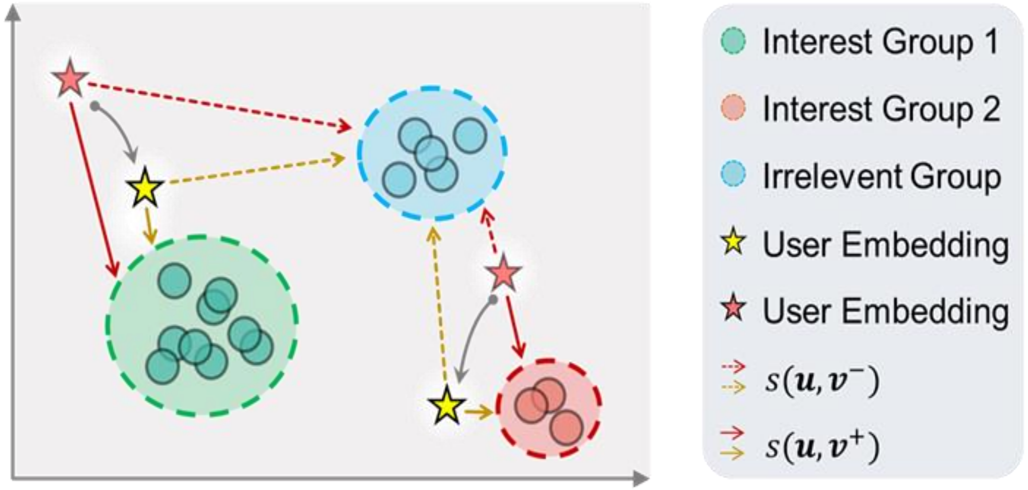
User Embeddings

Item Embeddings

- Diversity Control Regularization Scheme (DCRS)

$$\delta_{\boldsymbol{g},u_i} = \frac{1}{2C(C-1)} \sum_{c_1,c_2 \in [C]} \|\boldsymbol{g}_{u_i}^{c_1} - \boldsymbol{g}_{u_i}^{c_2}\|^2$$

$$\psi_{\boldsymbol{g}}(u_i) = \max(0, \delta_1 - \delta_{\boldsymbol{g},u_i}) + \max(0, \delta_{\boldsymbol{g},u_i} - \delta_2)$$



A proper value $\delta_{\boldsymbol{g},u_i}$ is significant to enjoy
a good generalization performance

Methodology

NEURAL INFORMATION PROCESSING SYSTEMS

- Final optimization goal:

$$\min_{\boldsymbol{g}} \hat{\mathcal{L}}_{\mathcal{D}}(\boldsymbol{g}) := \hat{\mathcal{R}}_{\mathcal{D},\boldsymbol{g}} + \eta \cdot \hat{\Omega}_{\mathcal{D},\boldsymbol{g}}$$

Recommendation-related Risk

$$\hat{\mathcal{R}}_{\mathcal{D},\boldsymbol{g}} = \frac{1}{|\mathcal{U}|} \sum_{u_i \in \mathcal{U}} \frac{1}{n_i^+ n_i^-} \sum_{j=1}^{n_i^+} \sum_{k=1}^{n_i^-} \ell_g^{(i)}(v_j^+, v_k^-)$$

$$\ell_g^{(i)}(v_j^+, v_k^-) = \max(0, \lambda + s(u_i, v_j^+) - s(u_i, v_k^-))$$

DCRS-related Risk

$$\hat{\Omega}_{\mathcal{D},\boldsymbol{g}} = \frac{1}{|\mathcal{U}|} \sum_{u_i \in \mathcal{U}} \psi_{\boldsymbol{g}}(u_i)$$

$$\psi_{\boldsymbol{g}}(u_i) = \max(0, \delta_1 - \delta_{\boldsymbol{g},u_i}) + \max(0, \delta_{\boldsymbol{g},u_i} - \delta_2)$$

Question2: Could CML generalize well under the multi-vector representation strategy?

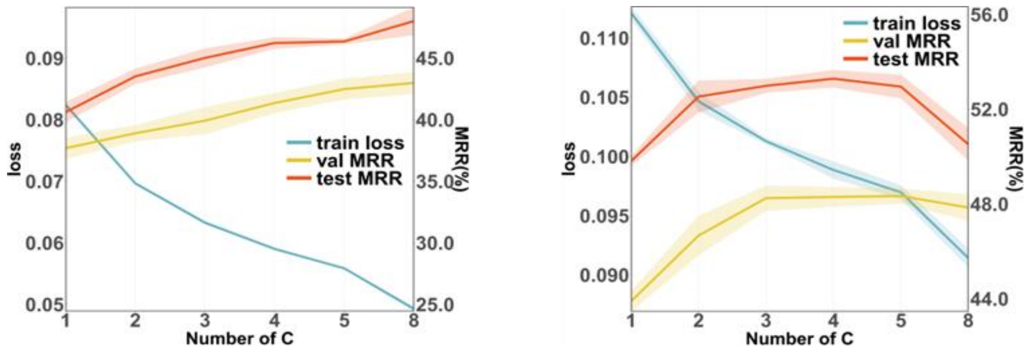
Methodology

Theorem 1. Generalization Upper Bound of CML

Let $\mathbb{E}[\hat{\mathcal{L}}_{\mathcal{D}}(\boldsymbol{g})]$ be the population risk of $\hat{\mathcal{L}}_{\mathcal{D}}(\boldsymbol{g})$. Then, $\forall\ \boldsymbol{g} \in \mathcal{H}_R$, with high probability, the following inequation holds:

Population Risk

$$\mathbb{E}[\hat{\mathcal{L}}_{\mathcal{D}}(\boldsymbol{g})] \leq \hat{\mathcal{L}}_{\mathcal{D}}(\boldsymbol{g}) + \sqrt{\frac{2d\log\left(3r\tilde{N}\right)}{\tilde{N}}}$$



DPCML could achieve a **smaller empirical risk** (**related to vector number C**)

It depends on the hypothesis space and data (**not related to C**)

DPCML could enjoy a smaller generalization error than CML

Experiments

NEURAL INFORMATION PROCESSING SYSTEMS

Basic Information of the Datasets. %Density is defined as $\frac{\#Ratings}{\#Users \times \#Items} \times 100\%$.

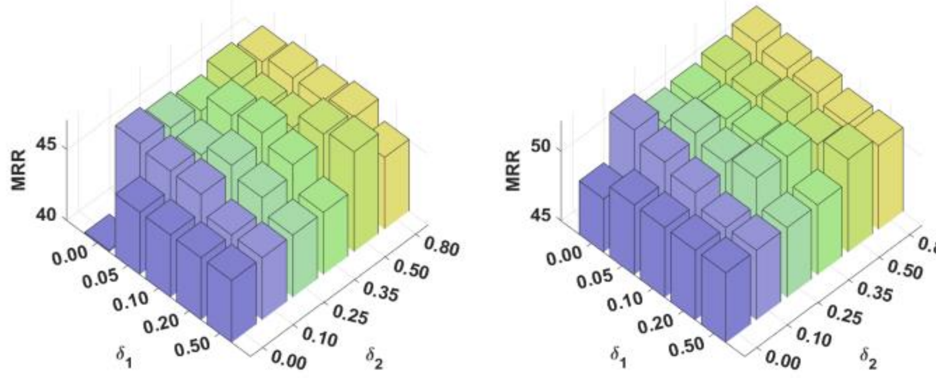
Datasets	MovieLens-1M	Steam-200k	CiteULike-T	MovieLens-10M
Domain	Movie	Game	Paper	Movie
#Users	6,034	3,757	5,219	69,167
#Items	3,953	5,113	25,975	10,019
#Ratings	575,271	115,139	125,580	5,003,437
%Density	2.4118%	0.5994%	0.0926%	0.7220%

- We perform the empirical studies over 4 benchmark datasets.
- We evaluate our proposed DPCML method against 14 state of the art competitors.

- We evaluate the performance with 8 metrics to show the Top-K recommendation performance.
- We also evaluate the diversity performance with the max-sum diversification metric.

Experiments

Dataset	Type	Method	P@3	R@3	NDCG@3	P@5	R@5	NDCG@5	MAP	MRR
MovieLens-1m	Item-based	itemKNN	12.24	2.90	12.41	12.43	4.29	12.79	8.34	26.16
	MF-based	GMF	14.10	2.81	14.33	14.28	4.08	14.73	8.29	29.51
		MLP	13.95	2.78	14.22	14.06	3.98	14.56	8.30	29.39
		NeuMF	16.43	3.20	16.87	16.73	4.68	17.40	9.69	33.23
		M2F	8.61	1.84	9.36	7.60	2.30	8.67	2.95	20.40
		MGMF	17.38	3.51	18.08	17.63	5.05	18.52	10.12	35.15
	CML-based	UniS	17.56	3.71	17.89	18.34	5.60	18.79	12.40	35.77
		PopS	12.96	3.11	13.30	12.82	4.41	13.40	7.59	28.61
		2stS	21.07	4.84	21.35	21.81	7.07	22.29	14.42	40.36
		HarS	24.88	5.86	25.38	24.89	8.25	25.77	15.74	45.15
		TransCF	10.03	1.84	10.31	10.90	3.09	11.20	7.07	23.66
		LRML	17.15	3.52	17.56	17.45	5.12	18.08	10.42	34.36
		AdaCML	19.06	4.12	19.31	19.74	6.23	20.20	13.30	37.36
		HLR	21.10	4.80	21.53	21.61	7.06	22.28	13.95	40.71
	Ours	DPCML1	19.12	4.14	19.34	19.90	6.27	20.29	13.24	37.55
		DPCML2	**25.18**	**6.06**	**25.64**	**25.35**	**8.51**	**26.16**	**16.09**	**45.32**
MovieLens-10m	Item-based	itemKNN	11.44	3.70	11.78	12.27	4.93	12.63	8.25	25.85
	MF-based	GMF	13.55	3.87	13.91	14.67	5.41	15.13	9.14	28.91
		MLP	15.27	4.93	15.46	16.08	6.53	16.38	12.77	32.21
		NeuMF	15.19	5.02	15.27	16.09	6.65	16.24	12.76	31.87
		M2F	7.03	1.41	7.21	7.55	2.23	7.98	2.50	15.17
		MGMF	14.62	4.26	15.15	15.53	5.96	16.26	10.30	31.07
	CML-based	UniS	10.15	2.84	10.33	11.19	4.08	11.38	8.92	24.24
		PopS	8.61	3.06	8.96	8.34	3.76	8.84	6.08	20.97
		2stS	16.47	4.89	16.72	17.62	6.87	18.06	12.89	33.75
		HarS	17.00	4.97	17.16	18.34	6.96	18.70	13.14	34.20
		TransCF	11.00	3.70	10.91	11.62	4.94	11.61	7.99	23.67
		LRML	13.72	3.96	13.98	14.53	5.58	15.08	8.99	28.77
		AdaCML	13.65	4.00	13.82	14.64	5.52	14.98	11.13	29.58
		HLR	15.13	5.12	14.94	16.40	7.00	16.23	13.40	31.66
	Ours	DPCML1	12.73	3.82	13.05	13.12	5.07	13.72	10.32	28.65
		DPCML2	**18.00**	**5.46**	**18.37**	**18.97**	**7.37**	**19.57**	**14.01**	**36.44**



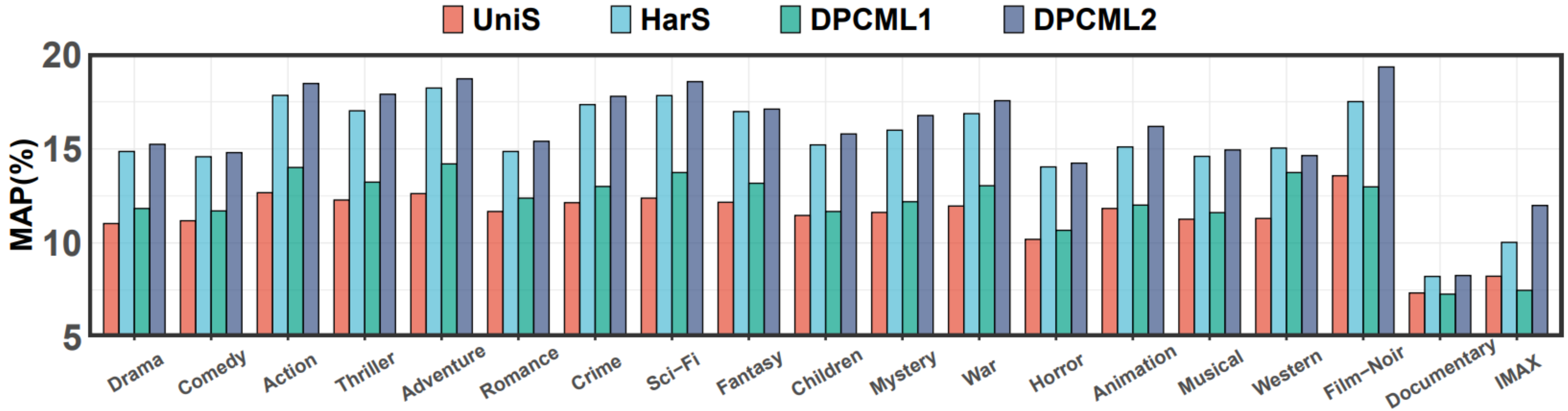
(a) DPCML1 (MRR) (b) DPCML2 (MRR)

- The experimental results show that our proposed algorithm DPCML could consistently outperform the state-of-the-art recommendation approaches.

Experiments

- DPCML could not only improve the recommendation performance but also promote the recommendation diversity.

Method	MaxDiv@3	MaxDiv@5	MaxDiv@10	MaxDiv@20
	Steam-200k			
UniS	1.354	4.750	23.520	117.927
HarS	1.752	6.809	40.378	236.794
DPCML1 w/o DCRS	1.643	5.857	30.425	155.193
DPCML1	1.822	6.713	34.727	179.065
DPCML2 w/o DCRS	2.958	11.398	65.398	365.458
DPCML2	2.977	11.472	65.952	369.876



Conclusion

Algorithm Level	We propose a novel algorithm called Diversity-Promoting Collaborative Metric Learning (DPCML) to accommodate the diversity of user preferences.
Algorithm Level	A novel diversity control regularization scheme (DCRS) is specifically tailored to explore the diverse interest of users better.
Theoretical Level	We start an early trial to present the generalization ability of the DPCML, and the results show that DPCML could generalize well to unseen test data.

Thank You!
Questions?

GitHub

WeChat