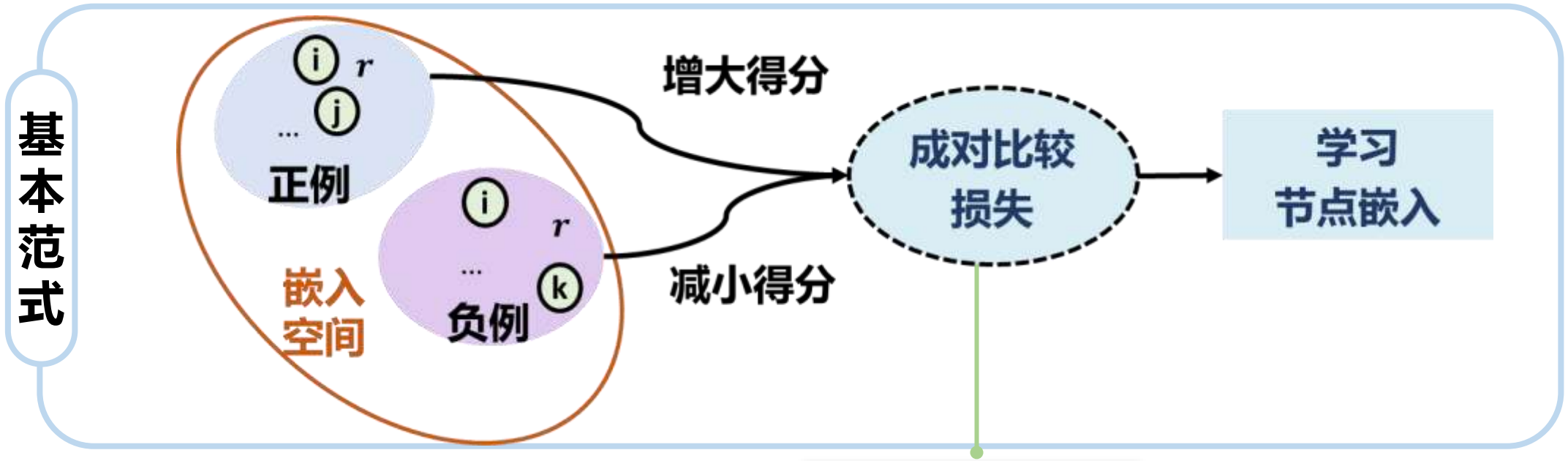
高效可泛化的鲁棒协同排序学习

包世龙
中国科学院大学 计算机科学与技术学院
baoshilong@ucas.ac.cn
2025年8月24日

总体研究背景

回顾：协同排序学习范式



正样本数

负样本数

基本损失函数形式

$$\min_f \quad \hat{\mathcal{R}}_{\mathcal{S}}(f) = \frac{1}{n^+ n^-} \sum_{j=1}^{n^+} \sum_{k=1}^{n^-} \ell\left(v_j^+, v_k^-\right)$$

Pairwise Loss

$$\ell\left(v_j^+, v_k^-\right) = \max(0, \lambda + \boldsymbol{d}(i,j) - \boldsymbol{d}(i,k))$$

Similarity Function

核心思路：学习候选样本间的相对优劣关系

网络空间中的协同排序学习典型场景

针对用户偏好优化
个性化推荐列表

捕捉模态间相关性
提升排序效果

结合任务上下文
排序候选结果

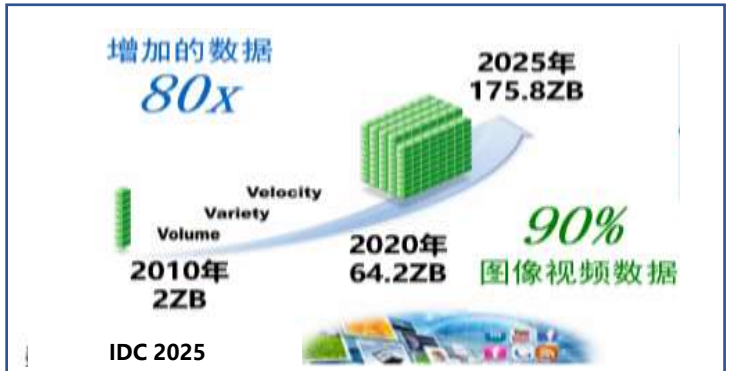
根据病历实现可疑病
灶区域定位

在多种机器学习下游任务应用广泛

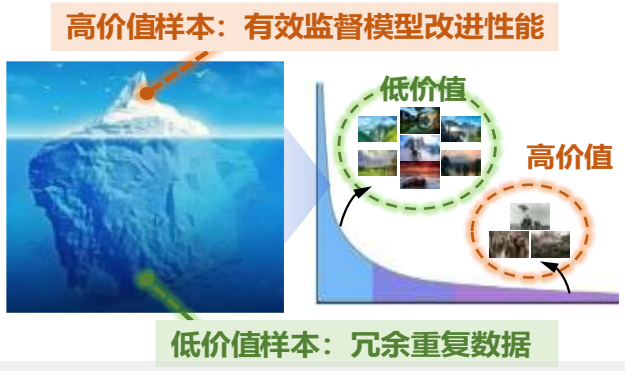
研究现状与挑战

聚焦网络空间中机器学习任务的数据特点

- “**数据规模庞大、样本价值稀疏、对抗攻击频发**”的网络空间任务特点，给协同排序学习的**泛化性和鲁棒性**带来巨大挑战



数据规模庞大



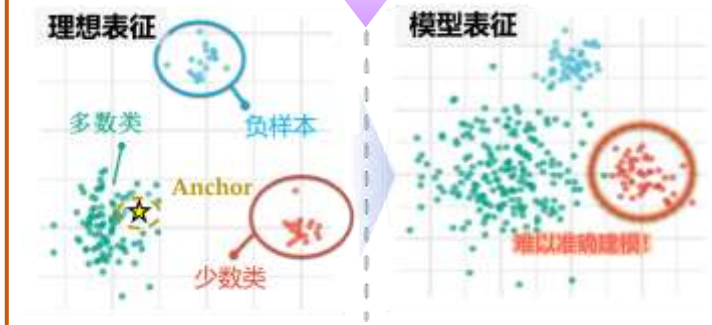
样本价值稀疏

对抗攻击频发

平方级别复杂度！

$$\hat{\mathcal{R}}_S(f) = \frac{1}{n^+ n^-} \sum_{j=1}^{n^+} \sum_{k=1}^{n^-} \ell\left(v_j^+, v_k^-\right)$$

样本耦合计算效率低



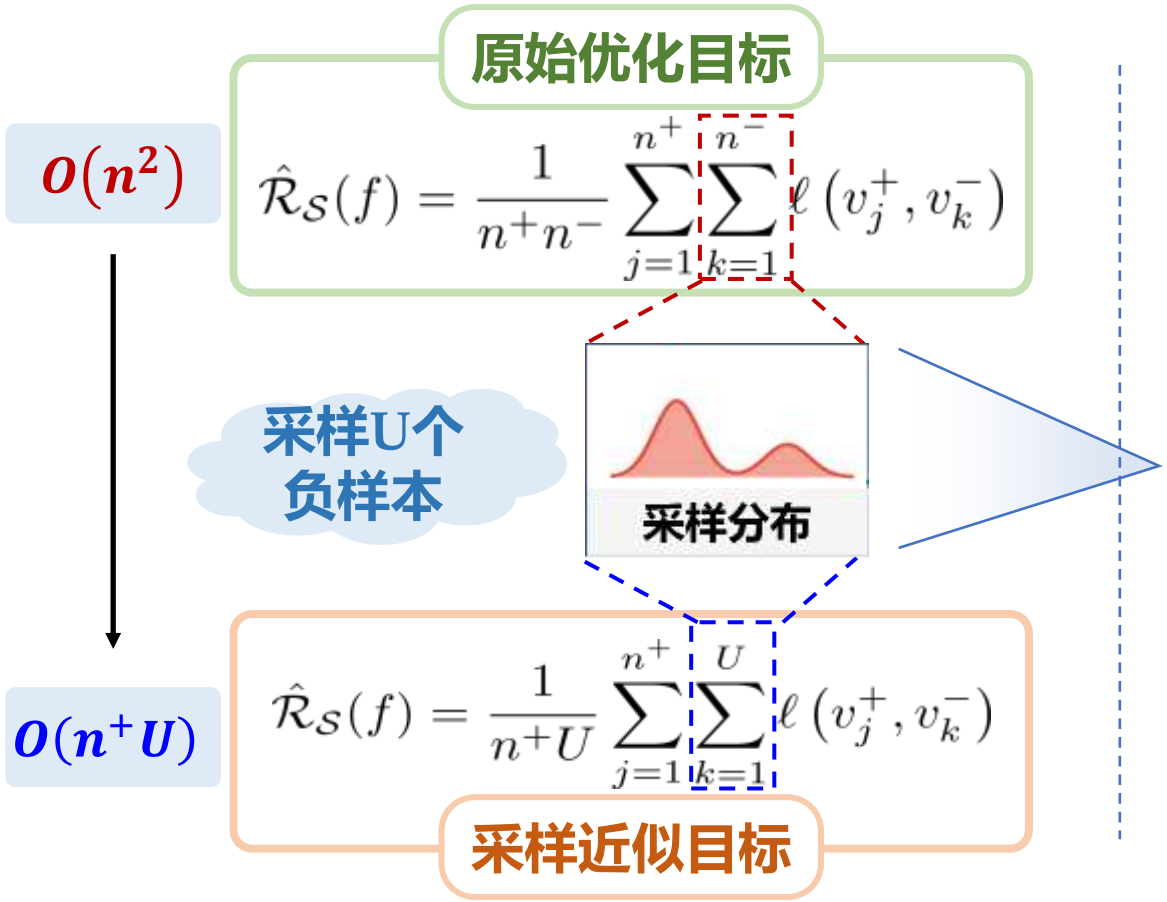
数据长尾表征偏差大

对抗扰动模型容错差

挑战1：样本耦合计算效率低

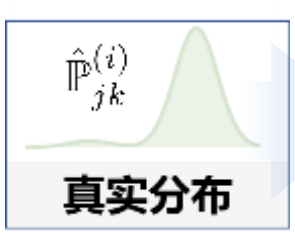
口基本思路：在训练时随机**采样部分负例**进行训练，从而简化计算

提高训练效率



- ✓ 均匀分布采样 [Pan 2008; C.-K. Hsieh 2017]
- ✓ 流行度/频率采样 [T. Chen 2017; G. Wu 2019]
- ✓ 对抗采样 [Cai 2017; V.-A. Tran 2019; Chen 2021]
- ✓ 难样本采样 [Ying 2018; J. Ding 2019; Shi 2023]
- ✓ 混合采样 [Yang 2022; Shi 2024]
- ✓ ……

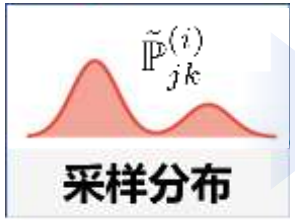
挑战1：样本耦合计算效率低



原始优化目标

$$\hat{\mathcal{R}}_{\mathcal{S}}(f) = \frac{1}{n^+ n^-} \sum_{j=1}^{n^+} \sum_{k=1}^{n^-} \ell\left(v_j^+, v_k^-\right)$$

负采样改变了原始优化目标，
致使泛化性能不明确



$$\hat{\mathcal{R}}_{\mathcal{S}}(f) = \frac{1}{n^+ U} \sum_{j=1}^{n^+} \sum_{k=1}^{U} \ell\left(v_j^+, v_k^-\right)$$

采样近似目标

现有方法仅局限于经验层面分析，
缺乏对相关问题的理论探究

f_{BGD}: Learning Embeddings From Positive Unlabeled Data with BGD

UAI'2018

Fajie Yuan,[1] Xin Xin,[1] Xiangnan He,[2] Guibing Guo,[3] Weinan Zhang,[4]
Tat-Seng Chua[2] and Joemon M. Jose[1]
University of Glasgow, UK[1], National University of Singapore, Singapore[2]
Northeastern University, China[3], Shanghai Jiao Tong University, China[4].
{f.yuan.1,x.xin.1,Joemon.Jose}@research.gla.ac.uk , xiangnanhe@gmail.com

AAAI'2020 Efficient Heterogeneous Collaborative Filtering without Negative Sampling for Recommendation

Chong Chen,[1] Min Zhang,[1] Yongfeng Zhang,[2] Weizhi Ma,[1] Yiqun Liu,[1] Shaoping Ma[1]
[1]Department of Computer Science and Technology, Institute for Artificial Intelligence,
Beijing National Research Center for Information Science and Technology, Tsinghua University
[2]Department of Computer Science, Rutgers University
cc17@mails.tsinghua.edu.cn, z-m@tsinghua.edu.cn

TKDE'2023 Graph-Based Non-Sampling for Knowledge Graph Enhanced Recommendation

Shuang Liang, Jie Shao, Jiasheng Zhang, and Bin Cui

Sampling a fraction of non-observed data as negative may **ignore other useful examples**, and thus lead to **non-optimal** performance.

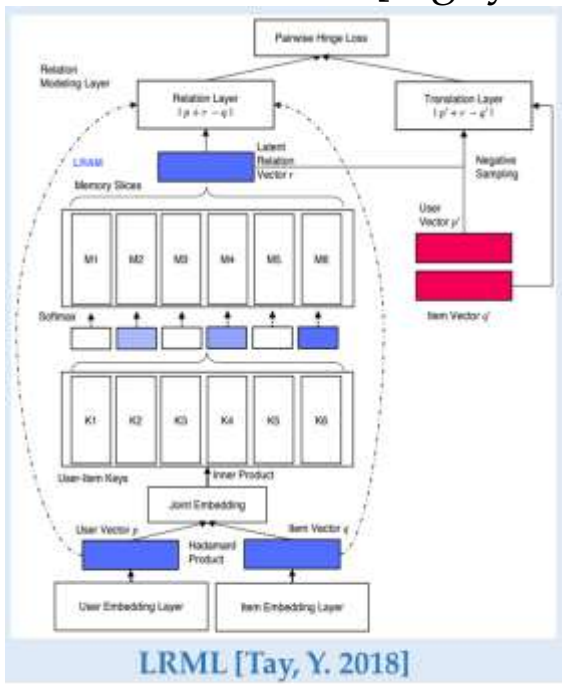
• • •

Sampling is **not robust and biased**, making it **difficult to converge** to the optimal ranking performance.

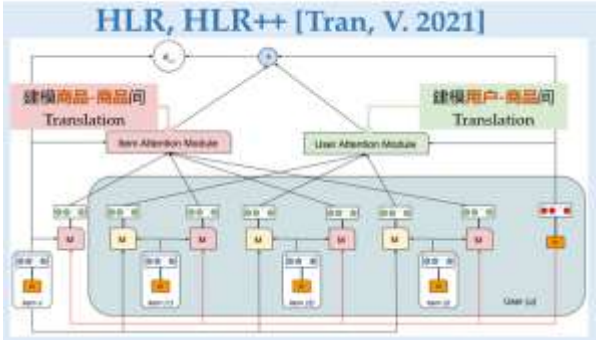
挑战2：数据长尾表征偏差大

口基本思路：通过改进模型结构**提高**对复杂数据的**建模能力**

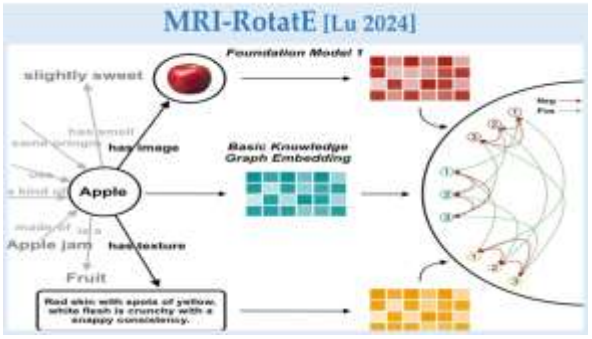
- ✓ TransCF [Park, C. 2017]; LRML [Tay, Y. 2018]; AdaCML [Zhang 2019]
- ✓ LightGCN [He 2020]; HLR, HLR++ [Tran, V. 2021]; IPML-PA [Zheng 2023];
- ✓ COMAUC [Nguyen 2023]; PBR [Zhao 2025]; ……



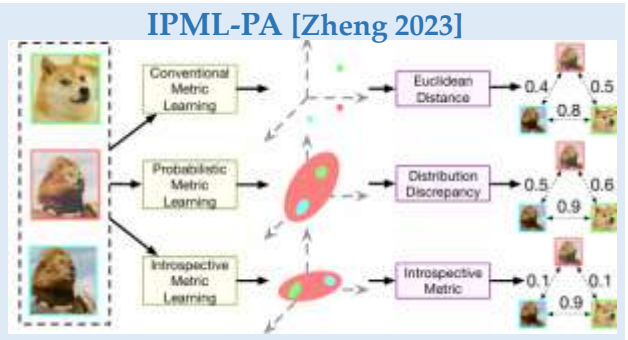
LRML [Tay, Y. 2018]



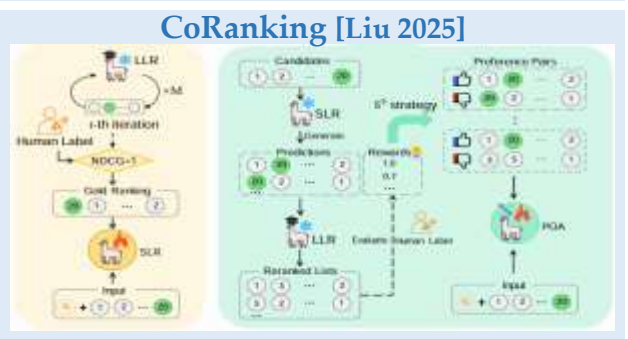
HLR, HLR++ [Tran, V. 2021]



MRI-RotatE [Lu 2024]



IPML-PA [Zheng 2023]



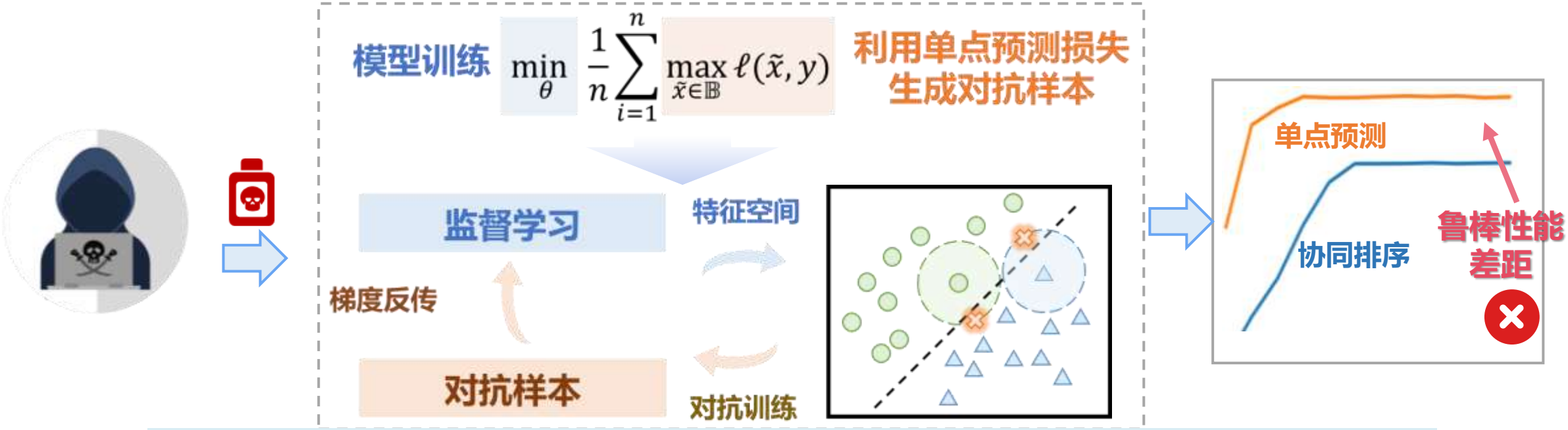
CoRanking [Liu 2025]

强表征≠强泛化，二者关系缺乏理论支撑，模型有效性边界不清晰

挑战3：对抗扰动模型容错差

口基本思路：主流方法大多**聚焦单点预测的鲁棒性**，借助对抗训练、正则化等进行算法设计与优化

- ✓ 对抗训练 [Kurakin, C. 2017]; [Madry 2018]; [Zhang 2019];
- ✓ Top-k CR [Jia 2020]; R2ET [Chen 2023]; RobRank [Zhou 2024];



缺乏对协同排序对抗鲁棒性的直接刻画，模型鲁棒性边界不明确

总体研究内容

总体研究内容

通过协同排序学习泛化理论分析结果，指导高效可泛化的鲁棒算法设计与优化

研究目标

高效可泛化的鲁棒协同排序学习

技术思路

构建协同排序学习的泛化理论分析框架

研究挑战

样本耦合	数据长尾	对抗扰动
计算效率低	表征偏差大	模型容错差

研究内容

高效无采样的协同排序学习方法	多表征平衡的协同排序学习方法	可证明鲁棒的协同排序学习方法
• 线性复杂度的全样本加速机制 • 理论层面验证方法泛化性能	• 表征与泛化性能间的对应关系 • 多向量自适应表征学习策略	• 保持协同排序鲁棒性的边界条件 • 端到端可证明的鲁棒排序学习
TPAMI 23 (一作); TPAMI 21/23; ICML 24/25 (Spotlight);	TPAMI 24 (一作); NeurIPS 22/24; NeurIPS 22 (一作, Oral, 1.7%)	TPAMI 25 (一作); TPAMI 23; ICML 21/22 (Long Talk);

协同排序学习的泛化理论分析框架

- Rademacher 复杂度是衡量假设类$\mathcal{F}$在给定样本上拟合随机噪声能力，具有**数据依赖性、可计算性，且复杂度越小，泛化越好**
- 传统泛化界依赖于对给定假设空间的 Rademacher 复杂度分析:

交叉熵损失

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \ell\left(\tilde{f}_{\theta}\left(x_i\right), y_i\right)$$

$$\hat{\mathfrak{R}}_S(\mathcal{G}) = \mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\sup_{g \in \mathcal{G}} \frac{1}{m} \sum_{i=1}^{m} \sigma_i g\left(z_i\right)\right]$$

$$\mathop{\mathbb{E}}_{S}\left[\sup_{f \in \mathcal{H}} \mathop{\mathbb{E}}_{S}\left(\hat{\mathcal{R}}_S\right) - \hat{\mathcal{R}}_S\right]$$

互相独立项的总和

样本不独立
不适用!

$$\hat{\mathcal{R}}_S(f) = \frac{1}{n^+ n^-} \sum_{j=1}^{n^+} \sum_{k=1}^{n^-} \ell\left(v_j^+, v_k^-\right)$$

互相不独立项的总和

协同排序学习难以直接套用传统框架，相关泛化理论的研究相对空白

协同排序学习的泛化理论分析框架

口理论层面—拓展传统基于 Rademacer 复杂度的泛化理论分析框架

定义 1. 协同排序学习的 Rademacher 复杂度

协同排序学习的经验 Rademacher 复杂度定义如下：

$$\hat{\mathfrak{R}}_{\ell,\mathcal{S}}^{cml}\left(\mathcal{H}_R\right)=\frac{1}{M}\sum_{u_i\in\mathcal{U}}\mathop{\mathbb{E}}_{\sigma_i}\left[\sup_{\mathcal{H}_R}\frac{1}{n_i^+n_i^-}\sum_{j=1}^{n_i^+}\sum_{k=1}^{n_i^-}\mathcal{Q}_{(i)}^{jk}\right]$$

其中

逐对样本

$$\mathcal{Q}_{(i)}^{jk}=\frac{\sigma_{ij}^+ + \sigma_{ik}^-}{2}\cdot\ell^{(i)}\left(v_j^+,v_k^-\right);$$

$\hat{\mathfrak{R}}_S(G)=\mathrm{E}_{\boldsymbol{\sigma}}\left[\sup_{g\in G}\frac{1}{m}\sum_{i=1}^{m}\sigma_i g\left(z_i\right)\right]$

单个样本

$\boldsymbol{\sigma}_i=\left[\sigma_{i1}^+,\sigma_{i2}^+,\ldots,\sigma_{in_i^+}^+,\sigma_{i1}^-,\sigma_{i2}^-,\ldots,\sigma_{in_i^-}^-\right]$ 是独立同分布的Rademacher随机变量。相应地，协同排序学习的期望Rademacher复杂数为：

$$\mathfrak{R}_{\ell,\mathcal{S}}^{cml}\left(\mathcal{H}_R\right)=\mathop{\mathbb{E}}_{\mathcal{S}}\left[\hat{\mathfrak{R}}_{\ell,\mathcal{S}}^{cml}\left(\mathcal{H}_R\right)\right]$$

实现成对比较损失的逐对样本解耦，克服泛化理论分析难的问题

协同排序学习的泛化理论分析框架

□理论层面—泛化理论结果

定理 2. 协同排序学习的泛化误差上界

对于任意$\delta \in (0,1)$, 至少$1-\delta$的概率有以下结论成立

$$\mathcal{R}_\ell^{cml}(f) \lesssim \hat{\mathcal{R}}_S^{cml}(f) + \phi \cdot \frac{\max(\lambda, \sqrt{R \cdot d})}{M} \cdot \sqrt{\frac{1}{\tilde{N}}} + \phi \cdot \frac{R}{M} \cdot \sqrt{\frac{\log 2/\delta}{2}} \cdot \sqrt{\frac{1}{\tilde{N}}}.$$

期望风险　　经验风险

有效样本量

$$\tilde{N} = \left(\sqrt{\frac{1}{n^+} + \frac{1}{n^-}} \right)^{-2}$$

定理 3. 基于采样的协同排序学习泛化误差上界

对于任意$\delta \in (0,1)$, 至少$1-\delta$的概率有以下结论成立

$$\mathcal{R}_\ell^{cml}(f) \lesssim \hat{\mathcal{R}}_S^{cml}(f) + \phi \cdot \frac{\max(\lambda, \sqrt{R \cdot d})}{M} \cdot \sqrt{\frac{1}{\tilde{N}}} + \phi \cdot \frac{R}{M} \cdot \sqrt{\frac{\log 2/\delta}{2}} \cdot \sqrt{\frac{1}{\tilde{N}}}$$

Lipschitz 常数项

$$+ \frac{(\lambda + 4R)}{M} \cdot \sum_{u_i \in \mathcal{U}} D_{TV}\left(\hat{\mathbb{P}}^{(i)}, \tilde{\mathbb{P}}^{(i)} \right)$$

采样分布与原分布间的**总变分距离**

支撑了协同排序学习的有效性，也在理论层面揭示了现有采样方法的局限性

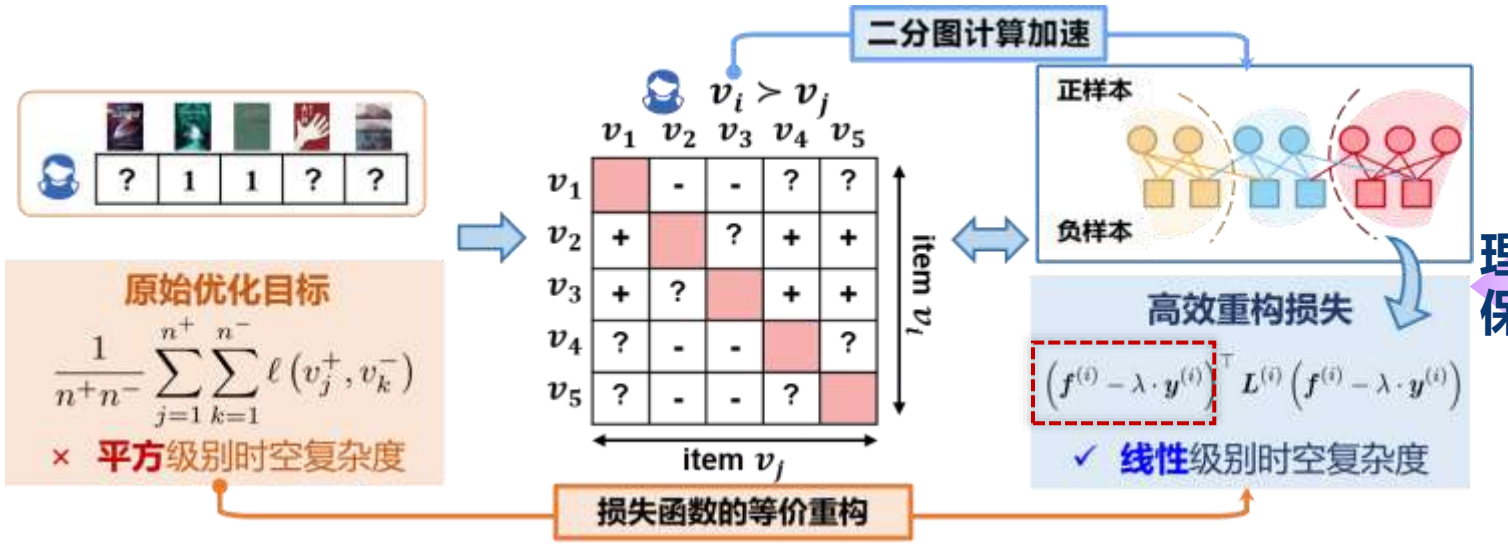
高效无采样的协同排序学习方法

样本耦合
计算效率低

口主要贡献—无需采样的全样本端到端高效优化

✓ 对原始优化目标进行等价损失重构，并进一步利用**二分图加速损失、梯度计算**

✓ 理论上证明所提方法**可消除**总变分距离的**采样偏差**，实现**合理泛化性能**



理论保障

命题. 线性时间复杂度可解

设 p 和 q 都是正整数, 对于任意矩阵 $P\in\mathbb{R}^{N\times p}$和 $Q\in\mathbb{R}^{N\times q}$，计算 $P^\top L\in\mathbb{R}^{p\times N}$ 可在 O(pN) 时间内完成，而 $P^\top LQ\in\mathbb{R}$ 则可以在 O(pqN)的时间内完成。

$P=Q, p=q=1$

定理 3. 具有同阶的泛化误差上界

对于任意$\delta\in(0,1)$,至少$1-\delta$的概率有以下结论成立

$$\mathcal{R}_\ell^{cml}(f)\lesssim\hat{\mathcal{R}}_S^{sfcml}(f)+(\lambda+4R)\cdot\frac{\max(\lambda,\sqrt{R\cdot d})}{M}\cdot\sqrt{\frac{1}{\tilde{N}}}+\frac{(\lambda+4R)\cdot R}{M}\cdot\sqrt{\frac{\log 2/\delta}{2}}\cdot\sqrt{\frac{1}{\tilde{N}}}$$

形成了高效可泛化的端到端优化方法，实现平方级别到线性复杂度的计算加速

高效无采样的协同排序学习方法

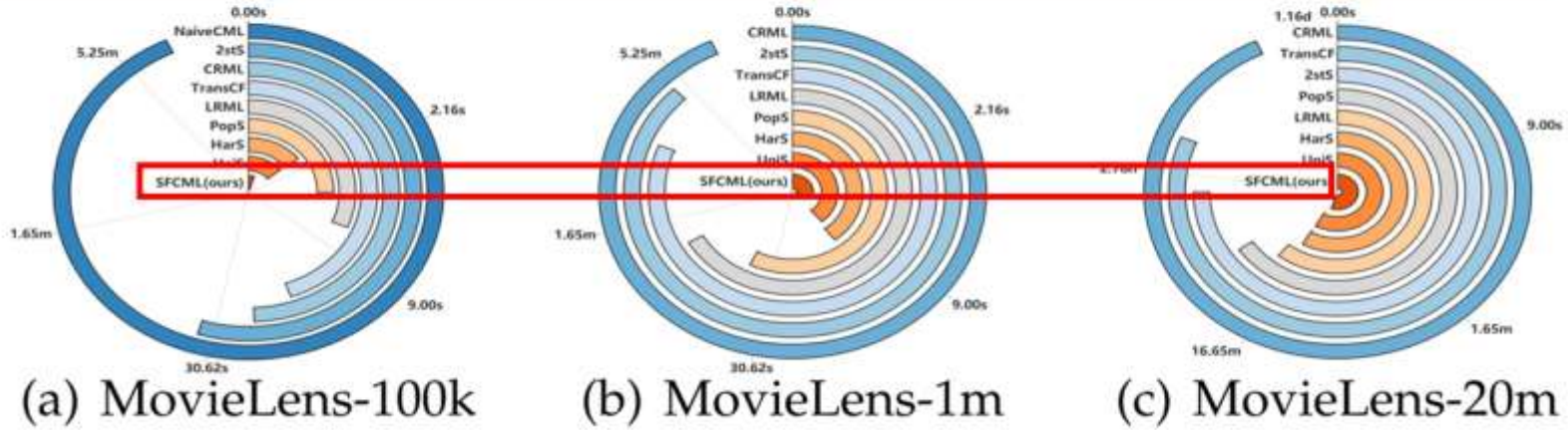
样本耦合
计算效率低

口实验结果

总体性能

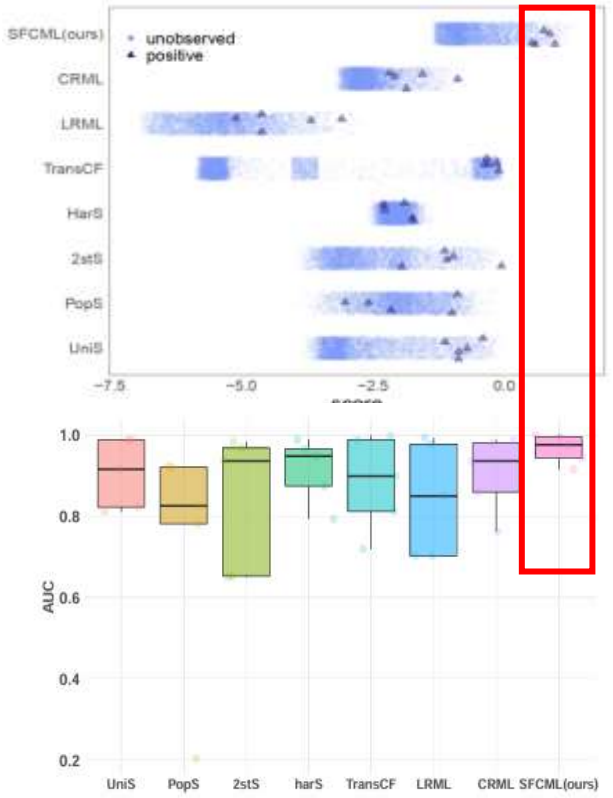
	Method	P@3	R@3	NDCG@3	P@5	R@5	NDCG@5	MAP	MRR	AUC
MovieLens-100k	itemKNN	11.35	2.41	11.57	12.96	4.11	13.45	8.49	24.63	85.68
	GMF	14.35	3.37	15.20	16.43	5.79	17.21	9.82	31.00	86.12
	MLP	14.98	3.93	15.57	15.51	5.70	16.54	10.09	31.99	87.09
	NCF	15.94	4.11	16.75	17.26	6.45	18.25	11.35	34.34	88.03
	EHCF	21.13	6.99	21.80	20.89	8.82	22.08	16.51	41.77	92.18
	UniS	15.94	4.43	16.06	17.04	6.23	17.40	13.21	33.07	92.27
	PopS	13.05	3.99	13.36	13.38	5.10	13.93	9.49	29.13	80.51
	2stS	15.50	4.42	15.77	16.76	6.21	17.18	13.35	32.95	92.01
	HarS	20.76	6.51	21.05	21.36	8.86	22.10	15.94	40.02	91.66
	TransCF	12.90	3.72	13.32	14.35	5.70	14.76	11.19	29.88	87.53
	LRML	20.65	6.65	21.44	20.36	8.24	21.75	13.48	37.93	90.38
	CRML	20.94	6.43	21.80	21.14	8.53	22.44	16.33	41.14	92.07
	NaiveCML	22.51	7.26	22.79	**23.85**	9.81	24.42	17.62	42.35	**93.24**
	SFCML(ours)	**23.40**	**7.62**	**23.63**	23.74	**9.95**	**24.65**	**18.00**	**43.13**	93.11

效率比较



(a) MovieLens-100k (b) MovieLens-1m (c) MovieLens-20m

头部样本排序性能



所提方法在多个数据集上达到**当前最佳性能**，**加速效率比达到2000+倍**

多表征平衡的协同排序学习方法

数据长尾
表征偏差大

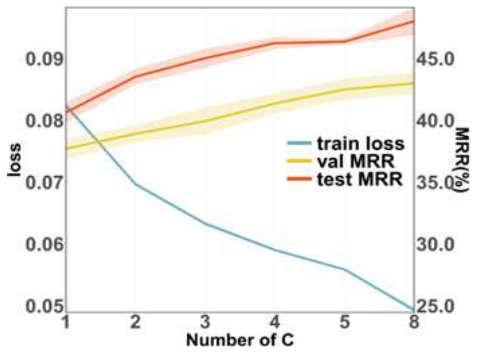
口理论层面

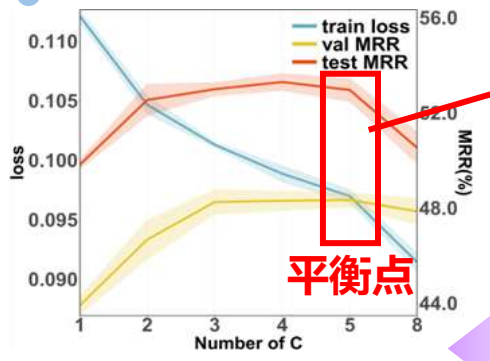
定理 4. 表征能力与泛化性能间对应关系 (Worst Case)

对于一般化的协同排序学习算法 ($C \geq 1$)，以较高的概率有如下结论成立

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

与C**无关**，取决于假设空间和数据集





需合理正则化控制！

增大向量个数C**不会显著增加模型的复杂度**，但会实现**更小的经验风险（更强的表征能力）**

理想情况

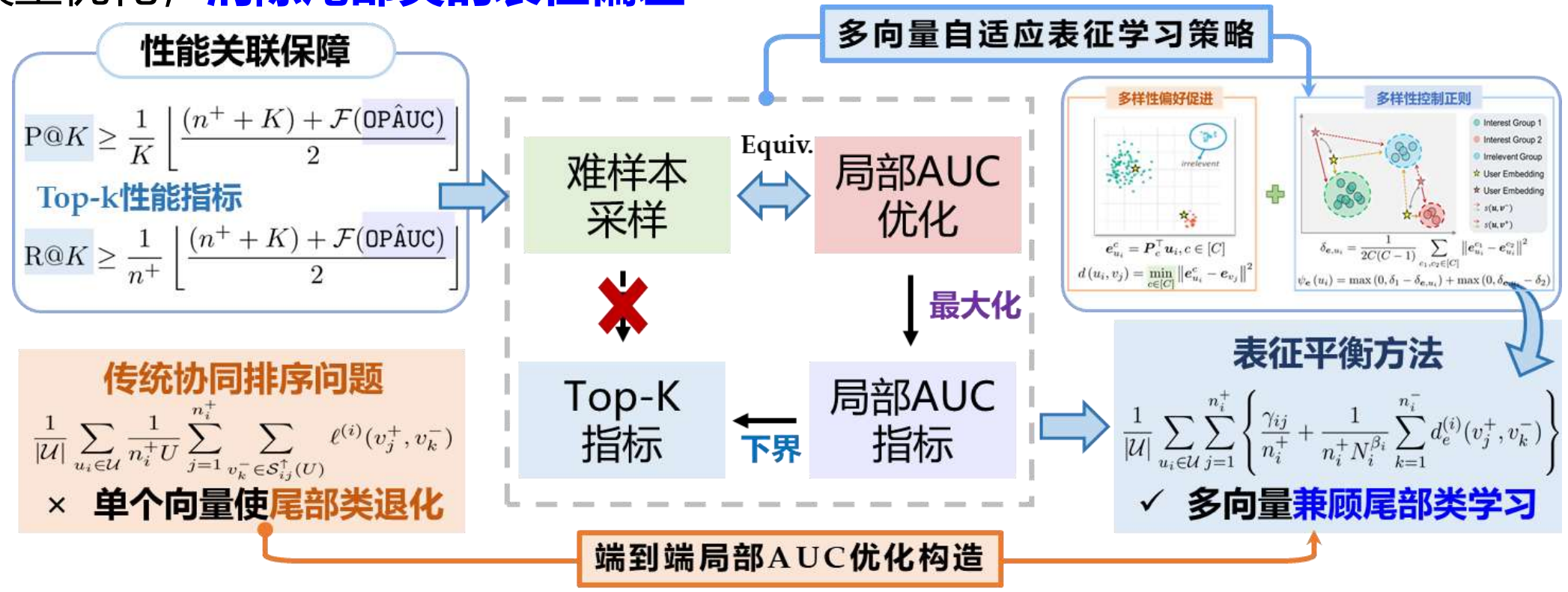
合理提升模型表征能力（C的个数）可提升模型的泛化性能

多表征平衡的协同排序学习方法

数据长尾
表征偏差大

口主要贡献

- ✓ 理论上得出表征能力与泛化性能对应关系，并构建**多向量自适应表征学习策略**
- ✓ 引入分布不敏感的 **局部AUC** 指标，利用其**与Top-K排序指标间的性能关联**指导模型优化，**消除尾部类的表征偏差**



多表征平衡的协同排序学习方法

数据长尾
表征偏差大

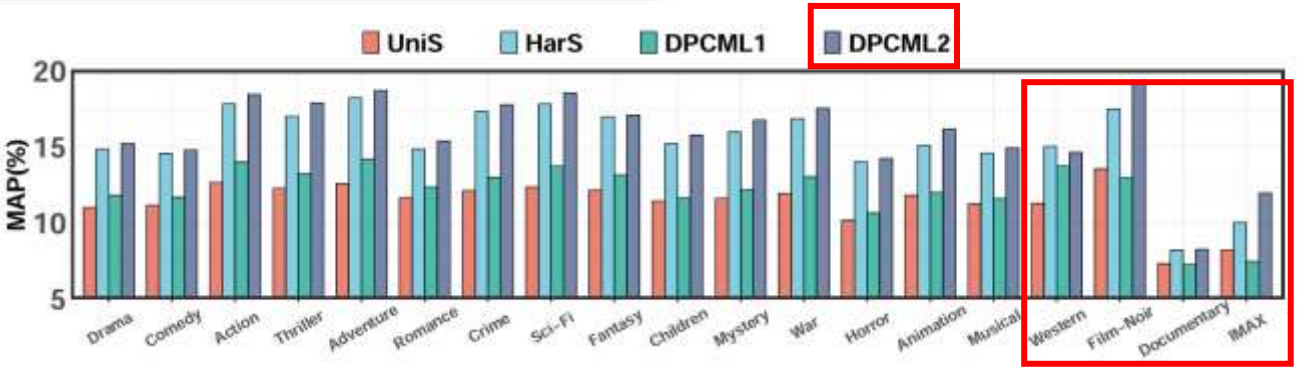
口实验结果

总体性能

	Type	Method	P@3	R@3	NDCG@3	P@5	R@5	NDCG@5	MAP	MRR
CiteULike	Item-based	itemKNN	1.20	0.83	1.23	1.15	0.77	1.16	1.44	3.78
	MF-based	BPR	6.47	3.50	6.84	7.89	4.05	8.49	5.14	16.20
		GMF	1.86	0.96	2.05	2.15	0.97	2.40	1.34	5.53
		MLP	2.06	1.08	2.22	2.40	1.16	2.61	1.52	12.37
		NeuMF	2.06	1.08	2.21	2.36	1.16	2.57	1.54	12.22
		M2F	1.76	0.90	1.97	1.87	0.93	2.18	0.93	4.53
		MGMF	2.31	1.23	2.48	2.42	1.12	2.71	1.51	6.18
	VAE-based	Mult-VAE	6.56	3.68	6.89	7.53	4.10	8.09	5.23	16.27
	GNN-based	LightGCN	8.33	4.64	8.68	9.58	5.23	10.23	6.32	19.14
	CML-based	UniS	7.34	3.71	7.48	9.54	5.13	10.02	5.59	17.27
		PopS	5.41	2.94	5.77	6.75	3.62	7.23	4.61	14.39
		2st	6.40	3.35	6.77	8.27	4.29	8.81	4.99	15.87
		HarS	8.44	4.41	8.82	10.43	5.60	11.25	6.67	20.08
		LRML	2.52	1.33	2.58	3.06	1.64	3.19	1.91	6.45
		TransCF	5.79	3.03	6.09	7.45	3.93	7.84	4.54	14.50
		AdaCML	7.04	3.75	7.31	8.70	4.52	9.18	5.57	17.31
		HLR	2.03	1.08	2.20	2.25	1.13	2.52	1.45	5.86
	DPCML-based	BPA+UniS	7.78	4.04	8.14	10.03	5.33	10.64	6.08	18.75
		APA+UniS	7.99	4.17	8.36	10.00	5.23	10.69	6.08	19.03
		BPA+HarS	8.70	4.59	9.06	10.96	5.85	11.47	6.44	19.96
		APA+HarS	8.82	4.73	9.18	11.02	5.87	11.56	6.68	20.30
		BPA+DiHarS	9.05	4.76	9.45	10.73	5.66	11.58	6.53	20.32
		APA+DiHarS	**9.24**	**4.94**	**9.72**	**11.20**	**5.99**	**12.09**	**6.72**	**20.88**

推荐冷启动问题

Type	Method	WarmStart			ColdStart User			ColdStart Item		
		P@3	R@3	N@3	P@3	R@3	N@3	P@3	R@3	N@3
Subset 1										
Joint-Training	MGMF+DN	11.55	4.56	11.56	2.47	1.07	2.26	9.05	3.26	9.16
	CML+DN	11.63	4.59	11.56	3.56	1.44	3.42	13.69	4.62	13.37
	DPCML+DN	12.27	4.89	12.21	7.26	3.15	7.08	14.79	**5.27**	14.57
Pre-Training	MGMF+DN	11.34	4.41	11.42	6.33	2.83	6.62	4.64	1.81	4.37
	CML+DN	11.87	4.66	11.90	6.49	2.64	5.74	14.57	4.98	**15.27**
	DPCML+DN	**12.60**	**4.99**	**12.67**	**8.65**	**3.27**	**8.41**	**15.45**	5.15	15.04
Subset 2										
Joint-Training	MGMF+DN	27.50	12.76	27.76	10.84	3.41	7.65	6.56	3.84	9.13
	CML+DN	27.78	12.88	27.86	23.44	**7.38**	21.81	8.51	4.66	9.34
	DPCML+DN	27.52	12.69	27.63	**24.90**	7.34	**26.31**	17.31	9.08	17.34
Pre-Training	MGMF+DN	27.98	12.98	28.11	12.35	3.93	9.03	14.14	8.08	15.71
	CML+DN	28.51	13.13	28.55	12.70	3.91	12.43	19.65	10.33	20.17
	DPCML+DN	**29.07**	**13.40**	**29.17**	14.88	3.90	15.43	**20.18**	**11.73**	**20.23**



尾部类性能

在多个数据集上达到**最佳性能**，并显著**提升尾部类性能**

可证明鲁棒的协同排序学习方法

对抗扰动
模型容错差

口协同排序学习的对抗鲁棒性问题

定义 2. Pairwise Adversarial Robustness

给定深度模型$h: \chi \to [0,1]$，对于任意一对样本(x_i^+, x_j^-)，若$h(x_i^+) > h(x_j^-)$，则对抗扰动后仍有以下关系成立：

$$h(x_i^+ + \delta_i) > h(x_j^- + \delta_j); \quad ||\delta_i||, ||\delta_j|| \leq \epsilon$$

扰动前后得分相对关系

对抗扰动项

对抗训练

FGSM [Goodfellow 15]、PGD [Madry 2018]

A Naïve Solution ✖

Minimax 对抗训练

$$\min_{\theta} \max_{\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_n} \sum_{i=1}^{n^+} \sum_{j=1}^{n^-} \frac{\ell\left[h_{\boldsymbol{\theta}}\left(\boldsymbol{x}_i^+ + \boldsymbol{\delta}_i\right) - h_{\boldsymbol{\theta}}\left(\boldsymbol{x}_j^- + \boldsymbol{\delta}_j\right)\right]}{n_+ n_-}$$

✖ (L1) 对抗样本求解开销大，难以直接优化！
(L2) 损失依赖性，仅抵御特定攻击
(L3) ……

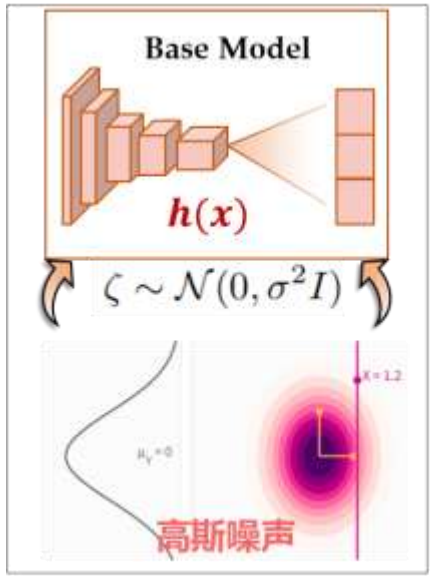
转而探索具有**形式化理论边界**的可证明鲁棒性方法

可证明鲁棒的协同排序学习方法

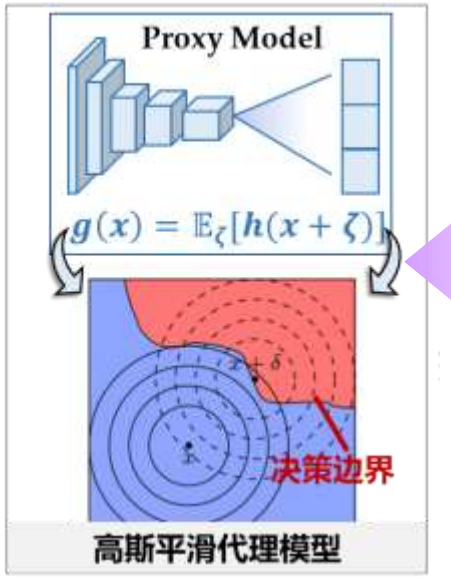
对抗扰动
模型容错差

口主要贡献—将随机平滑方法引入协同排序框架

✓ 基本思路：先**对输入样本添加高斯噪声**，**再**利用基础分类器**预测**，构建一个“平滑分类器”，通过平滑操作提升模型在对抗扰动下的鲁棒性



随机平滑



理论

引理 1

设 $\sigma > 0$，函数 $h : \mathcal{X} \to [0,1]$ 是可测的，定义：

$$g(x) = \mathbb{E}_{\zeta \sim \mathcal{N}(0,\sigma^2 I)}[h(x+\zeta)]$$

那么函数 $g(x)$ 关于 x 是 $\sqrt{\frac{2}{\pi\sigma^2}}$-Lipschitz.

引理 2

设 $\sigma > 0$，函数 $h : \mathcal{X} \to [0,1]$ 是可测的，定义：

$$g(x) = \mathbb{E}_{\zeta \sim \mathcal{N}(0,\sigma^2 I)}[h(x+\zeta)]$$

设 $\Phi(\cdot)$ 是标准高斯分布的累积分布函数（c.d.f.），则 $\Phi^{-1}(g(x))$ 关于 x 是 $\frac{1}{\sigma}$-Lipschitz.

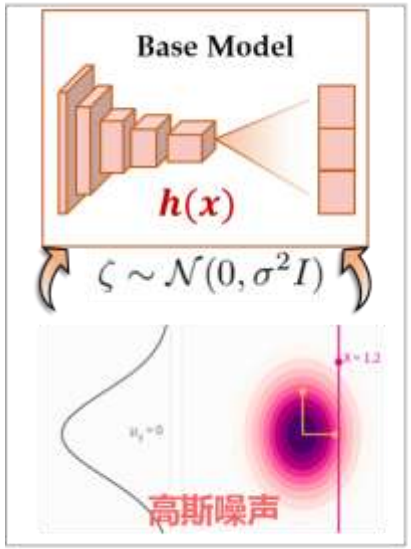
具有良好的理论性质，且易于扩展至任意深度的模型中来

可证明鲁棒的协同排序学习方法

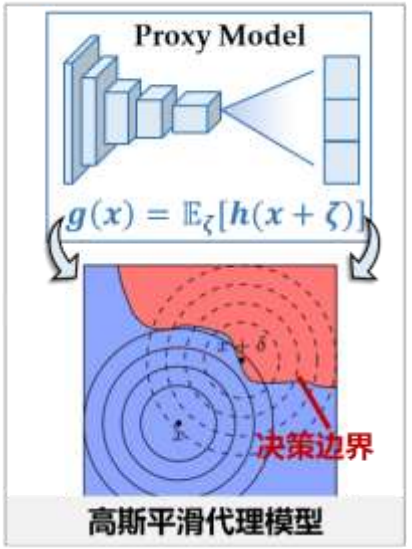
对抗扰动
模型容错差

口主要贡献—将随机平滑方法引入协同排序框架

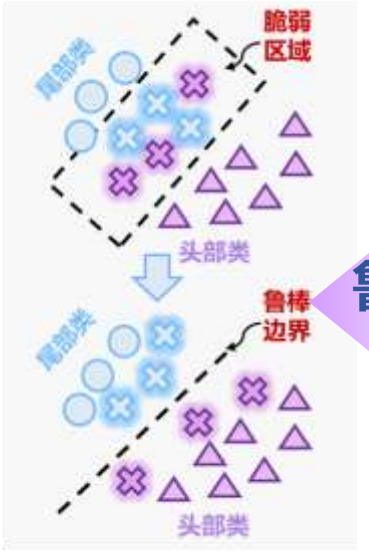
- ✓ 构建**面向协同排序模型的代理模型**近似原问题，并通过**最大化模型的潜在鲁棒安全区域**，确保模型抵御对抗攻击
- ✓ 理论上推导出模型的**鲁棒半径**及所提方法的**鲁棒泛化**误差上界



随机平滑



协同
排序学习



鲁棒性
保障

协同排序问题的鲁棒半径

给定任意协同排序模型$h: \chi \to [0,1]$及其高斯平滑代理模型g，对于任意一对样本(x_i^+, x_j^-),若满足$g(x_i^+) > g(x_j^-)$，若要使以下条件成立：

$$g(\tilde{x}_i^+) > g(\tilde{x}_j^-), \forall \tilde{x}_i^+ \in \mathcal{B}_{r_{ij}}(x_i^+), \tilde{x}_j^- \in \mathcal{B}_{r_{ij}}(x_j^-)$$

对抗样本

其最大扰动半径为

$$r_{ij} = \max\left\{\Delta_{ij}\sqrt{\frac{\pi\sigma^2}{8}}, \frac{\sigma}{2}\left(\Phi^{-1}(g(x_i^+)) - \Phi^{-1}(g(x_j^-))\right)\right\}$$

正负样本对得分之差　　标准高斯累积分布函数的逆

为鲁棒协同排序学习方法提供了理论保障与方法指导

可证明鲁棒的协同排序学习方法

对抗扰动
模型容错差

口理论层面

✓ 将可证明的鲁棒半径大小与潜在对抗攻击鲁棒性评估联系起来

Robustness Goal

(G1) $\mathop{\mathbb{E}}_{\substack{x^+\sim\mathcal{P},\\ x^-\sim\mathcal{N}}}\left[\ell_{0,1}^{adv}(x^+,x^-)\middle| g(x^+)>g(x^-)\right]=0$

损失尽可能小! 干净样本正确

$+\quad \mathcal{L}_{SG}=(G1)+(G2)$

干净样本上的性能尽可能高!

(G2) $\mathop{\mathbb{E}}_{\substack{x^+\sim\mathcal{P},\\ x^-\sim\mathcal{N}}}\left[\ell_{0,1}(g(x^+)-g(x^-))\right]=0$

Performance Goal

定理 5. (Informal) 攻击半径小于鲁棒半径

设 $\mathcal{R}_{\mathcal{X}}=\{r_{ij}\mid r_{ij}>0, x_i^+\in\mathcal{X}_P, x_j^-\in\mathcal{X}_N\}$ 是所有样本的鲁棒半径集合.
若对抗攻击最大扰动半径 $r_{\text{adv}}\le\min_{r_{ij}\in\mathcal{R}_{\mathcal{X}}} r_{ij}$，则以下结论以高概率成立:

$$\mathcal{L}_{\text{SG}}(g)\le \underbrace{\hat{\mathcal{L}}_D(g,\mathcal{X})}_{\text{经验风险}}+\sqrt{\frac{2A\log(C\tilde{N})}{\tilde{N}}}$$ 模型复杂度

定理 6. (Informal) 中高强度对抗攻击半径

记 $\rho(r_{\text{adv}})=\frac{\sum_{r\in\mathcal{R}_{\mathcal{X}}}\mathbb{I}[r<r_{\text{adv}}]}{|\mathcal{R}_{\mathcal{X}}|}$ 表示抵御 (ℓ_2,r_{adv}) 对抗攻击的失败率,
$\mathbb{C}_r$ 表示一对样本排序正确的概率. 当对抗攻击的最大扰动半径满足
$\min_{r_{ij}\in\mathcal{R}_{\mathcal{X}}} r_{ij}<r_{\text{adv}}\le\max_{r_{ij}\in\mathcal{R}_{\mathcal{X}}} r_{ij}$，则以下不等式以高概率成立:

$$\mathcal{L}_{\text{SG}}(g)\preceq\hat{\mathcal{L}}_{\text{D}}(g,\mathcal{X})+\sqrt{\frac{2}{\pi\sigma^2}r_{\text{adv}}\left(1+\hat{\mathbb{C}}_r(\rho(r_{\text{adv}})-1)\right)}+\sqrt{\frac{2A\log(C\tilde{N})}{\tilde{N}}}$$

构建了可证明鲁棒性的泛化性能评估方法，并理论证明其有效性

可证明鲁棒的协同排序学习方法

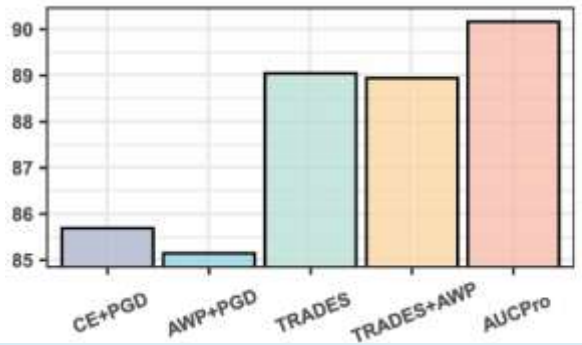
对抗扰动
模型容错差

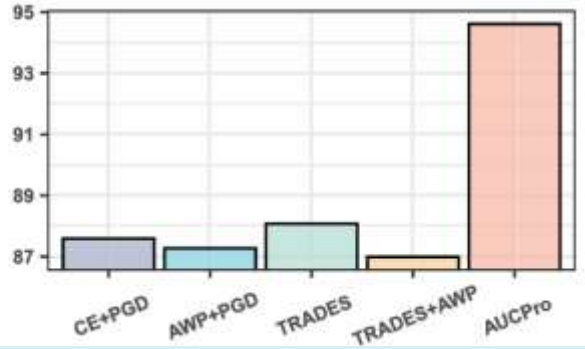
口实验部分

可证明鲁棒性

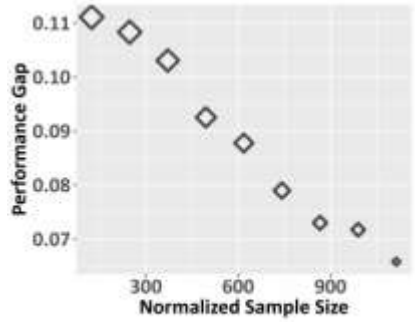
Methods	ACR-AUC	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
Subset 1														
Smooth [24]	2.69	98.72	97.32	95.98	94.29	92.15	89.55	86.22	81.92	76.35	69.35	60.90	51.40	41.35
Macer [92]	1.42	97.15	90.32	84.76	77.11	67.96	57.36	47.18	35.74	24.78	16.16	9.77	5.17	2.19
Consistency [67]	2.96	98.36	96.89	95.79	94.42	92.62	90.33	87.36	83.62	78.93	73.30	66.74	59.35	51.48
AUCPro (Ours)	3.11	98.04	96.12	95.00	93.67	91.97	89.84	87.11	83.67	79.46	74.55	68.98	62.85	56.11
Subset 2														
Smooth [24]	2.56	99.75	99.22	98.22	96.42	93.42	89.41	84.38	78.10	70.72	62.42	53.67	44.76	35.91
Macer [92]	1.81	98.91	97.35	94.44	90.14	83.08	72.45	59.37	46.16	34.01	24.37	17.95	13.01	9.34
Consistency [67]	3.25	99.82	99.60	99.15	98.30	97.08	95.57	93.39	90.42	86.43	81.26	74.91	67.75	59.99
AUCPro (Ours)	3.84	99.24	98.21	97.39	96.29	95.32	94.44	93.36	92.00	90.24	88.09	85.39	81.97	77.92
Subset 3														
Smooth [24]	1.38	94.94	89.27	84.19	77.45	68.21	56.25	44.01	32.84	22.96	15.07	9.28	5.42	2.97
Macer [92]	1.23	93.55	88.92	82.93	73.93	62.41	49.62	36.01	25.13	15.52	7.83	3.26	1.00	0.23
Consistency [67]	1.91	94.21	91.58	88.51	84.61	79.66	73.42	66.12	58.03	49.06	39.58	30.24	21.88	14.95
AUCPro (Ours)	2.26	95.58	91.60	89.36	86.75	83.60	79.75	74.97	69.05	62.10	54.31	45.90	37.33	28.91

对抗防御

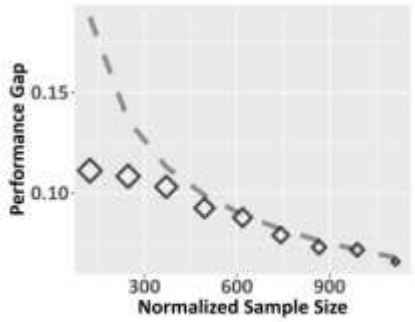




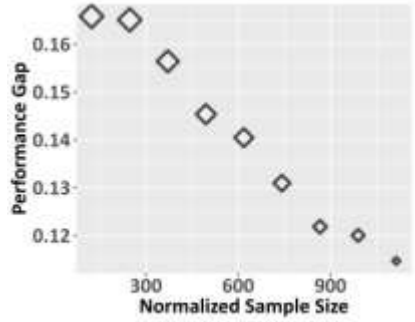
鲁棒泛化界的经验验证



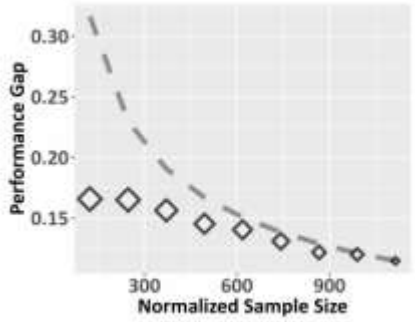
(a) $r_{adv} = 0.00$



(b) A=0.369, $r_{adv} = 0.00$



(c) $r_{adv} = 1.00$



(d) A=1.054, $r_{adv} = 1.00$

在多种对抗攻击场景下取得了当前最佳性能

论文发表

- **Shilong Bao,** Qianqian Xu, Zhiyong Yang, Yuan He, Xiaochun Cao, and Qingming Huang. AUCPro: AUC-Oriented Provable Robustness Learning. **TPAMI, 2025** (IF=18.6, **CCF-A**).
- **Shilong Bao,** Qianqian Xu, Zhiyong Yang, Yuan He, Xiaochun Cao, and Qingming Huang. Improved Diversity-Promoting Collaborative Metric Learning for Recommendation. **TPAMI, 2024** (IF=18.6, **CCF-A**).
- **Shilong Bao**, Qianqian Xu, Zhiyong Yang, Xiaochun Cao and Qingming Huang. Rethinking Collaborative Metric Learning: Toward an Efficient Alternative without Negative Sampling. **TPAMI, 2023** (IF=18.6, **CCF-A**).
- **Shilong Bao**, Qianqian Xu, Zhiyong Yang, Yuan He, Xiaochun Cao and Qingming Huang. The Minority Matters: A Diversity-Promoting Collaborative Metric Learning Algorithm. **NeurIPS, 2022** (Oral, 1.7%, **CCF-A**).
- **Shilong Bao**, Qianqian Xu, Ke Ma, Zhiyong Yang, Xiaochun Cao and Qingming Huang. Collaborative Preference Embedding against Sparse Labels. **ACM MM, 2019** (Oral, 5.6%, **CCF-A**).
- Cong Hua, Qianqian Xu, Zhiyong Yang, Zitai Wang, **Shilong Bao**, Qingming Huang. OpenworldAUC: Towards Unified Evaluation and Optimization for Open-world Prompt Tuning. **ICML, 2025** (**CCF-A**).
- Feiran Li, Qianqian Xu, **Shilong Bao**, Zhiyong Yang, Runmin Cong, Xiaochun Cao, Qingming Huang. Size-invariance Matters: Rethinking Metrics and Losses for Imbalanced Multi-object Salient Object Detection. **ICML, 2024** (Spotlight, 3.5%, **CCF-A**).
- Boyu Han, Qianqian Xu, Zhiyong Yang, **Shilong Bao**, Peisong Wen, Yangbangyan Jiang and Qingming Huang. AUCSeg: AUC-oriented Pixel-level Long-tail Semantic Segmentation. **NeurIPS, 2024** (**CCF-A**).

论文发表

- Zhiyong Yang, Qianqian Xu, Wenzheng Hou, **Shilong Bao**, Yuan He, Xiaochun Cao and Qingming Huang. Revisiting AUC-oriented Adversarial Training with Loss-Agnostic Perturbations. **TPAMI, 2023** (**IF=18.6**, **CCF-A**).
- Zhiyong Yang, Qianqian Xu, **Shilong Bao**, Peisong Wen, Yuan He, Xiaochun Cao and Qingming Huang. AUC-Oriented Domain Adaptation: From Theory to Algorithm. **TPAMI, 2023** (**IF=18.6**, **CCF-A**).
- Wenzheng Hou, Qianqian Xu, Zhiyong Yang, **Shilong Bao**, Yuan He and Qingming Huang. AdAUC: End-to-end Adversarial AUC Optimization Against Long-tail Problems. **ICML, 2022** (**CCF-A**).
- Huiyang Shao, Qianqian Xu, Zhiyong Yang, **Shilong Bao** and Qingming Huang. Asymptotically Unbiased Instance-wise Regularized Partial AUC Optimization: Theory and Algorithm. **NeurIPS, 2022** (**CCF-A**).
- Zhiyong Yang, Qianqian Xu, **Shilong Bao**, Yuan He, Xiaochun Cao and Qingming Huang. When All We Need is a Piece of the Pie: A Generic Framework for Optimizing Two-way Partial AUC. **ICML, 2021** (**Long Talk, 3%**, **CCF-A**).
- Zhiyong Yang, Qianqian Xu, **Shilong Bao**, Xiaochun Cao and Qingming Huang. Learning with Multiclass AUC: Theory and Algorithms. **TPAMI, 2021** (**IF= 23.6**, **CCF-A**).

Thanks!

https://statusrank.github.io/

baoshilong@ucas.ac.cn