Learning Personalized Attribute Preference via Multi-task AUC Optimization

Zhiyong Yang,[1,2] Qianqian Xu,[3] Xiaochun Cao,[1,2] Qingming Huang [3,4]
[1]SKLOIS, IIE, CAS [2]School of Cyber Security, UCAS
[3]Key Lab of Intell. Info. Process., ICT, CAS
[4]UCAS

Introduction

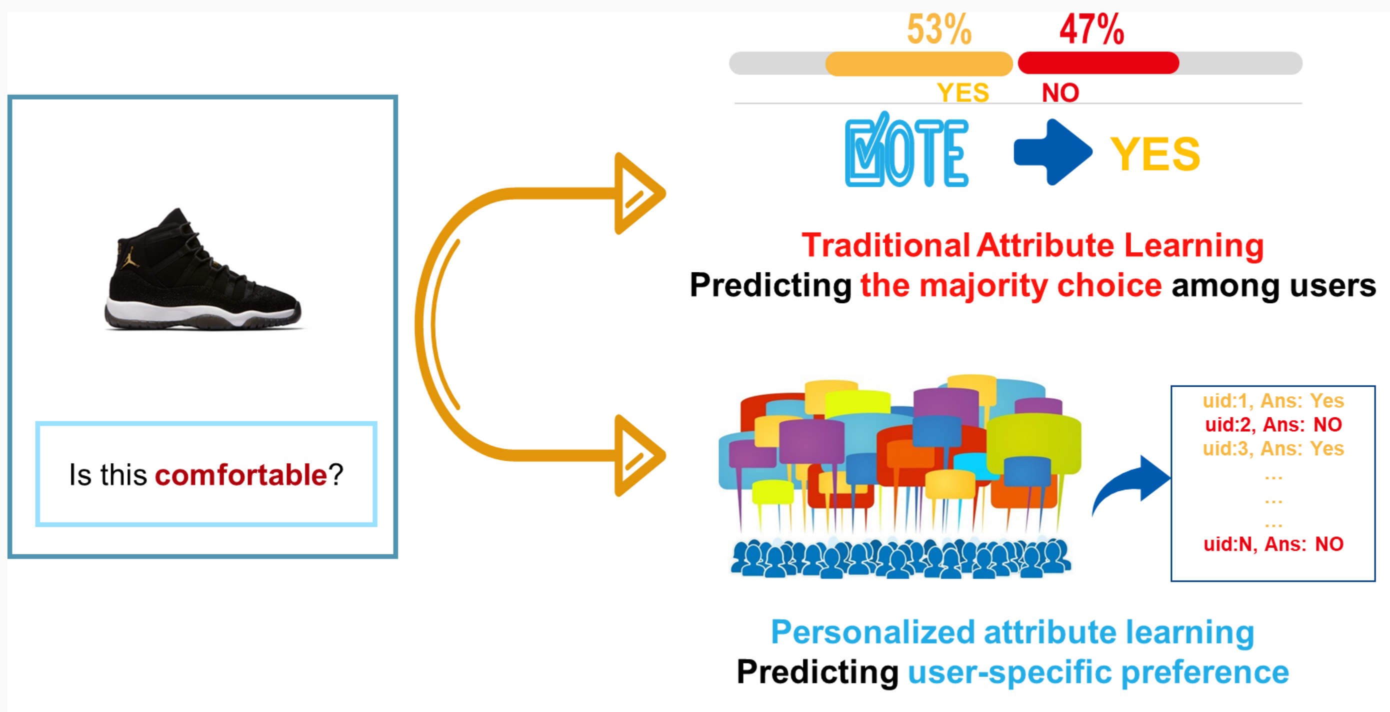


Fig. 1: illustration of our proposed model

Fig. 1 Instead of adopting a consensus learning framework, in our paper, we will propose effective solutions for personalized attribute learning.

In this paper, we are especially interested in the following two problems:

Problem A How to model the correlation of the user behaviors?

Problem B How to guarantee that a positive labeled instance has a higher rank than negatively labeled instances?

Model Formulation

- For a specific attribute, we are given user-specific labels from n_u different workers. Then the training data could be represented as:

$$\mathcal{T} = \left\{(\boldsymbol{X}^{(1)}, \boldsymbol{y}^{(1)}), \cdots, (\boldsymbol{X}^{(n_u)}, \boldsymbol{y}^{(n_u)})\right\}$$

- Our goal is to learn a predictor $\boldsymbol{f}^{(i)}$ for each of the personalized label vectors $\boldsymbol{y}^{(i)}$. In this paper, we assume that $f^{(i)}(\cdot)$ has a linear form :

$$\boldsymbol{f}^{(i)} = \boldsymbol{X}^{(i)}\boldsymbol{w}^{(i)}$$

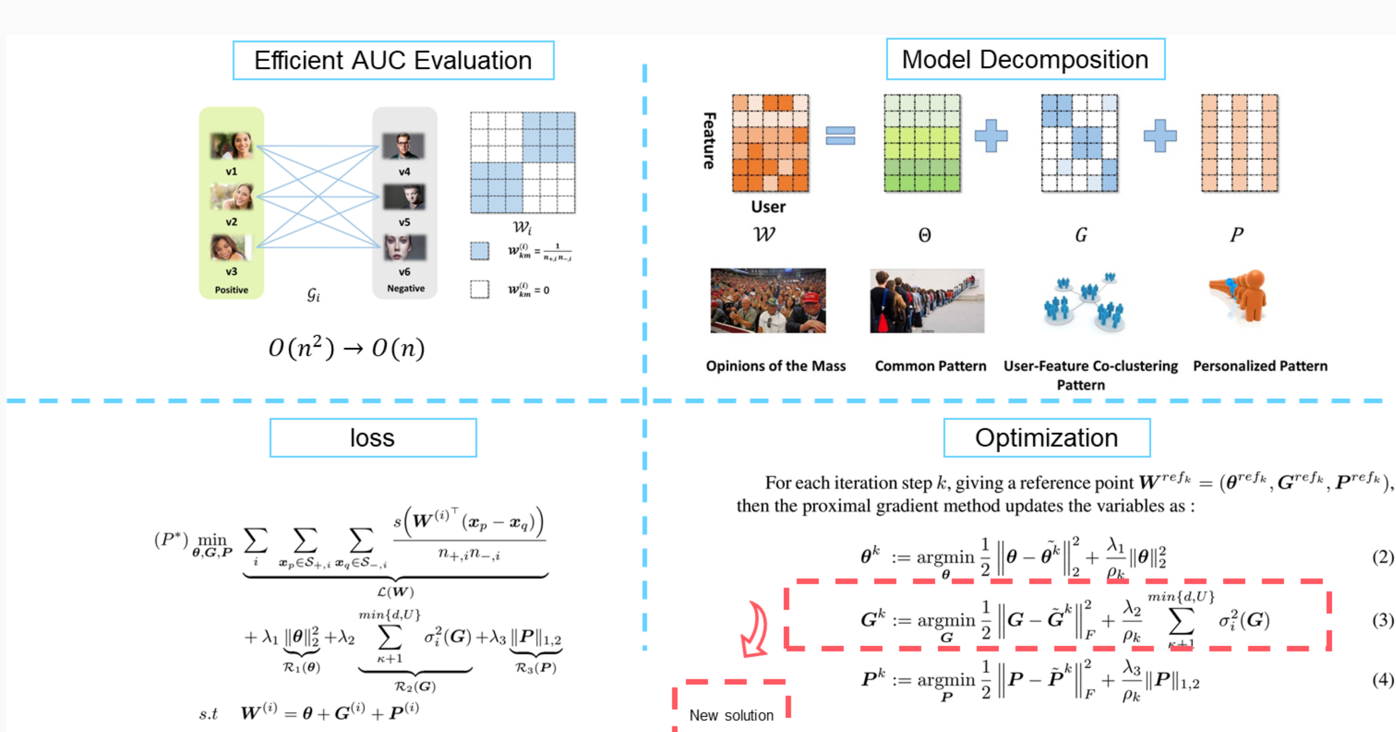
where $\boldsymbol{w}^{(i,j)}$ is the corresponding model weight.

- We adopt a three-component additive decomposition of $\boldsymbol{w}^{(i)}$:

$$\boldsymbol{w}^{(i)} = \boldsymbol{\theta} + \boldsymbol{G}^{(i)} + \boldsymbol{P}^{(i)} \quad (1)$$

- Objective function :

$$(P^*) \min_{\boldsymbol{\theta},\boldsymbol{G},\boldsymbol{P}} \underbrace{\sum_i \sum_{\boldsymbol{x}_p \in \mathcal{S}_{+,i}} \sum_{\boldsymbol{x}_q \in \mathcal{S}_{-,i}} \frac{s\left(\boldsymbol{W}^{(i)^\top}(\boldsymbol{x}_p - \boldsymbol{x}_q)\right)}{n_{+,i}n_{-,i}}}_{Problem\ A}$$
$$\underbrace{+ \lambda_1 \|\boldsymbol{\theta}\|_2^2 + \lambda_2 \sum_{\kappa+1}^{min\{d,U\}} \sigma_i^2(\boldsymbol{G}) + \lambda_3 \|\boldsymbol{P}\|_{1,2}}_{Problem\ B}$$
$$s.t \quad \boldsymbol{W}^{(i)} = \boldsymbol{\theta} + \boldsymbol{G}^{(i)} + \boldsymbol{P}^{(i)}$$



Theoretical Analysis

Theorem 1. • *1) The sequence $\{\mathcal{F}(\boldsymbol{\theta}^k, \boldsymbol{G}^k, \boldsymbol{P}^k)\}$ is non-increasing in the sense that :* $\forall k, \exists C_{k+1} > 0,$

$$\mathcal{F}(\boldsymbol{\theta}^{k+1}, \boldsymbol{G}^{k+1}, \boldsymbol{P}^{k+1}) \le \mathcal{F}(\boldsymbol{\theta}^k, \boldsymbol{G}^k, \boldsymbol{P}^k) - C_{k+1}(\left\|\Delta(\boldsymbol{\theta}^k)\right\|_2^2 + \left\|\Delta(\boldsymbol{G}^k)\right\|_F^2 + \left\|\Delta(\boldsymbol{P}^k)\right\|_F^2) \quad (2)$$

- *2)* $\lim_{k\to\infty} \boldsymbol{\theta}^k - \boldsymbol{\theta}^{k+1} = 0,\ \lim_{k\to\infty} \boldsymbol{G}^k - \boldsymbol{G}^{k+1} = 0,\ \lim_{k\to\infty} \boldsymbol{P}^k - \boldsymbol{P}^{k+1} = 0.$
- *3) The parameter sequences $\{\boldsymbol{\theta}^k\}_k$, $\{\boldsymbol{G}^k\}_k$, $\{\boldsymbol{P}^k\}_k$ are bounded*
- *4) Every limit point of $\{\boldsymbol{\theta}^k, \boldsymbol{G}^k, \boldsymbol{P}^k\}_k$ is a critical point of the problem.*
- *5)* $\forall\ T \ge 1, \exists\ C_T > 0$:

$$\min_{0\le k<T} \left(\left\|\Delta(\boldsymbol{\theta}^k)\right\|_2^2\right) \le \frac{C_T}{T},\ \min_{0\le k<T} \left(\left\|\Delta(\boldsymbol{G}^k)\right\|_F^2\right) \le \frac{C_T}{T},\ \min_{0\le k<T} \left(\left\|\Delta(\boldsymbol{P}^k)\right\|_F^2\right) \le \frac{C_T}{T}.$$

Theorem 2. *Assume that $\exists\Delta_\mathcal{X} > 0$, all the instances are sampled such that, $||x|| \le \Delta_\mathcal{X}$ .Define $C = (\psi_1 + \sqrt{\psi_2 + \kappa \cdot \sigma_{max}^2} + \psi_3)\ \zeta$ as $\zeta = \Delta_\mathcal{X} C$, we have, for all $\delta \in (0,1)$, for all $(\boldsymbol{\theta}, \boldsymbol{G}, \boldsymbol{P}) \in \Theta$:*

$$\mathbb{E}_\mathcal{D}(\sum_i \ell_{AUC}^{(i)}) \le \mathcal{L}(\boldsymbol{W}) + \sum_{i=1}^{U} \frac{B_1}{\sqrt{(n_i\chi_i(1-\chi_i))}} + B_2\sqrt{\frac{ln(\frac{2}{\delta})}{\sum_{i=1}^{U} n_i\chi_i(1-\chi_i)}}$$

holds with probability at least $1-\delta$, where $B_1 = 8\sqrt{2}C\Delta_\mathcal{X}(1+\zeta)$, $B_2 = 10\sqrt{2}(1+\zeta)\zeta$, $\chi_i = \frac{n_{+,i}}{n_i}$. The distribution $\mathcal{D} = \otimes_{i=1}^{U}(\mathcal{D}_{+,i} \otimes \mathcal{D}_{-,i})$, where for user i, $\mathcal{D}_{+,i}$, $\mathcal{D}_{-,i}$ are conditional distributions for positive and negative instances, respectively.

Experiments

Simulated Dataset

- We generate a annotation Dataset with 100 users with each user labeled 5000 samples.

Table 1 AUC Comparison on Simulation Dataset

Alg	RMTL	rMTFL	LASSO	JFL
mean	83.48	83.45	83.57	83.49
Alg	CMTL	COMT	RAMU	Ours
mean	83.47	83.44	83.50	**99.65**

Table 2 Running Time Comparison (seconds)

ratio	20%	40%	60%	80%	100%
Orginal	18.57	74.22	151.86	268.55	nan
Ours	**3.06**	**5.50**	**8.65**	**12.46**	**15.82**

Real-World Dataset

- **Shoes**: The Shoes Dataset contains 14,658 online shopping images. In this dataset, 7 attributes are annotated by users with a wide spectrum of interests and backgrounds. For each attribute, there are at least 190 users who take part in the annotation, and each user is assigned with 50 images. Overall, 90,000 annotations are collected in this dataset.
- **Sun**: The SUN Attributes Dataset is a well-known large-scale scene attribute dataset with roughly 1,4000 images and a taxonomy of 102 discriminative attributes. Recently, the personalized annotations over five attributes are collected with hundreds of annotators. For each person, 50 images are labeled based on their own comprehension and preference. Overall, this dataset contains 64,900 annotations collected from different users.

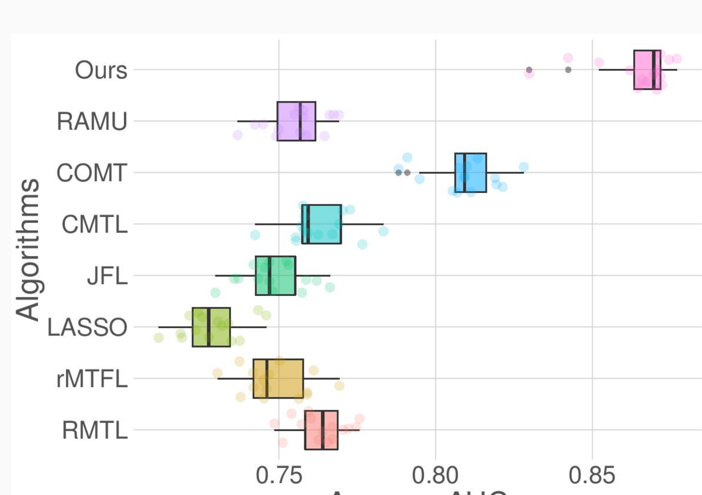


Fig. 3: Overall Performance on Shoes

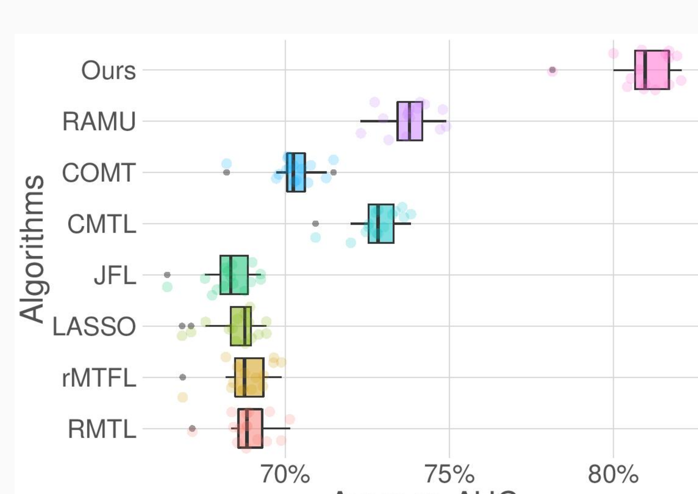


Fig. 4: Overall Performance on Shoes

Table 3 Performance Comparison based on the AUC metric

Alg	Attibutes											
	Shoes							Sun				
	BR	CM	FA	FM	OP	ON	PT	CL	MO	OP	RU	SO
RMTL	79.31	84.99	66.90	85.08	75.67	67.22	75.14	69.36	62.71	75.28	67.91	69.23
rMTFL	70.90	83.78	67.27	85.91	73.71	65.21	77.11	69.27	62.15	75.80	68.16	68.76
LASSO	68.46	80.48	65.90	84.01	71.47	64.60	75.08	67.64	61.83	75.39	68.57	69.13
JFL	72.00	83.10	67.26	85.93	73.02	65.39	77.09	68.63	61.94	75.00	67.17	68.78
CMTL	74.54	85.16	68.21	85.32	75.06	68.17	77.62	72.55	66.61	79.78	72.34	72.82
COMT	84.24	88.68	69.66	89.19	80.93	72.99	80.62	70.69	63.72	76.93	69.43	70.44
RAMU	78.33	84.58	65.78	84.68	75.25	66.72	73.50	72.95	69.25	79.81	74.39	72.50
Ours	**92.95**	**90.92**	**73.24**	**92.65**	**87.95**	**81.07**	**86.22**	**79.31**	**78.19**	**86.50**	**81.88**	**78.98**