When All We Need is a Piece of the Pie: A Generic Framework for Optimizing Two-way Partial AUC

Zhiyong Yang[1,2], Qianqian Xu[3], Shilong Bao[1,2], Yuan He[6], Xiaochun Cao[1,2], Qingming Huang[2,3,4,5*]

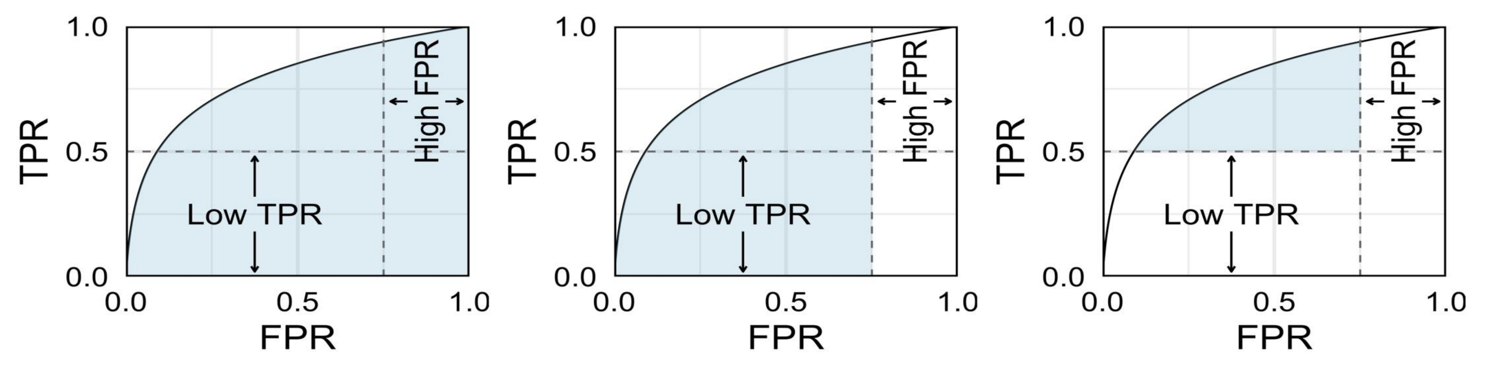
[1]Institute of Information Engineering, CAS [2]University of Chinese Academy of Sciences

[3]Institute of Computing Technology, CAS [4]BDKM, CAS [5]Peng Cheng Laboratory [6]Alibaba Group

Motivation

In most applications, a skillful classifier should simultaneously embrace a high TPR and a low FPR. However,

- Traditional AUC summarizes the average performance over all possible TPRs and FPRs, showing a biased estimation of the performance by involving some unrelated regions, Fig-(a).
- The One-way Partial AUC (OPAUC) only calculates the AUC within a pre-defined FPR range$[\alpha, \beta]$, Fig-(b).
- The Two-way Partial AUC (TPAUC) can serve this purpose shown in Fig-(c), which measures the area of a partial region of the ROC curve with $TPR \geq 1-\alpha, FPR \leq \beta$. But how to optimize the TPAUC metric directly has not yet been studied.



(a) AUC (b) OPAUC (c) TPAUC (our target)

Two-Way Partial AUC Metrics

- Define TPAUC as follows:

$$\text{TPAUC} = \int_{\xi}^{\beta} \text{TPR}\left(\text{FPR}^{-1}(\theta)\right) d\theta - (1-\alpha)\cdot(\beta-\xi)$$

FPR $\leq \beta$

TPR $\geq 1-\alpha$

$\xi = \text{FPR}\left(\text{TPR}^{-1}(1-\alpha)\right)$ → a function of the scoring function f

- Reformulation of TPAUC:

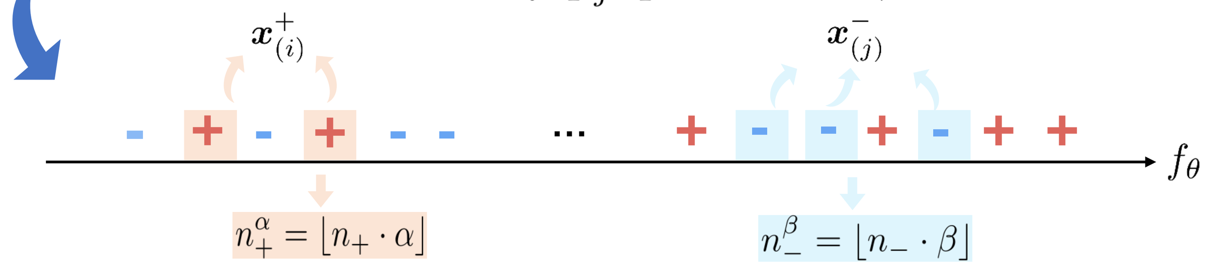
$$\text{AUC}_{\beta}^{\alpha}(f_\theta, \mathcal{S}) = \mathbb{E}\left[\mathbf{1}\left[f(\boldsymbol{x}) > f(\boldsymbol{x}'), f(\boldsymbol{x}) \leq t_{1-\alpha}, f(\boldsymbol{x}') \geq t_\beta | y=1, y'=-1\right]\right]$$

$$t_{1-\alpha} = \operatorname*{argmin}_{\delta \in \mathbb{R}} \left[\delta \in \mathbb{R}: \ \mathbb{E}\left[\mathbf{1}\left[f(\boldsymbol{x}^+) \leq \delta\right]\right] = \alpha\right]$$

$$t_\beta = \operatorname*{argmin}_{\delta \in \mathbb{R}} \left[\delta \in \mathbb{R}: \ \mathbb{E}\left[\mathbf{1}\left[f(\boldsymbol{x}^-) \geq \delta\right]\right] = \beta\right]$$

- Surrogate loss minimization: $\ell_{\exp}(t) = \exp(-t), \ell_{sq}(t) = (1-t)^2$

$$(OP_0) \min_{\theta} \hat{\mathcal{R}}^{\ell}_{\alpha,\beta}(S, f_{\boldsymbol{\theta}}) = \sum_{i=1}^{n_+^\alpha}\sum_{j=1}^{n_-^\beta} \frac{\ell\left(f_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{(i)}^+\right) - f\left(\boldsymbol{x}_{(j)}^-\right)\right)}{n_+^\alpha n_-^\beta}$$



Construct Surrogate Optimization Problems

- Bi-level optimization

$$(OP_1) \min_{\boldsymbol{\theta}} \frac{1}{n_+^\alpha n_-^\beta} \sum_{i=1}^{n_+}\sum_{j=1}^{n_-} v_i^+ \cdot v_j^- \cdot \ell\left(f_{\boldsymbol{\theta}}, \boldsymbol{x}_i^+, \boldsymbol{x}_j^-\right)$$

$$\text{s.t } v_+ = \operatorname*{argmax}_{v_i^+ \in [0,1]} \sum_{i=1}^{n^+} \left(v_i^+ \cdot \left(1 - f_{\boldsymbol{\theta}}\left(\boldsymbol{x}_i^+\right)\right) - \varphi_\gamma\left(v_i^+\right)\right)$$

$$v_- = \operatorname*{argmax}_{v_j^- \in [0,1]} \sum_{j=1}^{n_-} \left(v_j^- \cdot f_{\boldsymbol{\theta}}\left(\boldsymbol{x}_j^-\right) - \varphi_\gamma\left(v_j^-\right)\right)$$

Sample Weights

Penalty Function

The connection between weight and the penalty is the key!

- Dual correspondence

Definition 1 Calibrated Smooth Penalty Function regularities

A penalty function $\varphi_\gamma(x): \mathbb{R}_+ \to \mathbb{R}$*satisfies the following*

(A) φ_γ *has continuous third-order derivatives.*

(B) φ_γ *is strictly increasing in the sense that* $\varphi_\gamma'(x) > 0.$

(C) φ_γ *is strictly convex in the sense that* $\varphi_\gamma''(x) > 0.$

(D) φ_γ *has positive third-order derivatives in the sense that* $\varphi_\gamma'''(x) > 0.$

Definition 2 Calibrated Weighting Function

A weighting function $\psi_\gamma(x): [0,1] \to \mathrm{Rng}$ *, where* $\mathrm{Rng} \subseteq [0,1]$ *, satisfies the following regularities:*

(A) ψ_γ *has continuous second-order derivatives.*

(B) ψ_γ *is strictly increasing in the sense that* $\psi_\gamma'(x) > 0.$

(C) ψ_γ *is strictly concave in the sense that* $\psi_\gamma''(x) < 0.$

Proposition 1

Given a strictly convex function φ_γ *and define* $\psi_\gamma(t)$ *as*

$$\psi_\gamma(t) = \operatorname*{argmax}_{v \in [0,1]} \ v \cdot t - \varphi_\gamma(v)$$

Then we can draw the following conclusions:

(a) If φ_γ *is a calibrated smooth penalty function, we have* $\boxed{\psi_\gamma(t) = \varphi_\gamma'^{-1}(t).}$

penalty to weight

(b) If ψ_γ *is a calibrated weighting function such that* $v = \psi_\gamma(t)$ *we have* $\boxed{\varphi_\gamma(v) = \int \psi_\gamma^{-1}(v)dv + \text{const.}}$

weight to penalty

$$v_i^+ = \psi_\gamma\left(1 - f_{\boldsymbol{\theta}}\left(\boldsymbol{x}_i^+\right)\right), v_j^- = \psi_\gamma\left(f_{\boldsymbol{\theta}}\left(\boldsymbol{x}_j^-\right)\right), v_i^+, v_j^- \in [0,1]$$

If ψ_γ *has a closed-form expression*

Cancel the inner optimization problem

$$\hat{\mathcal{R}}^{\ell}_{\psi}(\mathcal{S}, f_{\boldsymbol{\theta}}) = \frac{1}{n_+ n_-} \sum_{i=1}^{n_+}\sum_{j=1}^{n_-} \psi_\gamma\left(1 - f_{\boldsymbol{\theta}}\left(\boldsymbol{x}_i^+\right)\right) \psi_\gamma\left(f_{\boldsymbol{\theta}}\left(\boldsymbol{x}_j^-\right)\right) \cdot \ell\left(f_{\boldsymbol{\theta}}, \boldsymbol{x}_i^+, \boldsymbol{x}_j^-\right)$$

- Theoretical analysis

Theorem 1 (Informal).

The following inequality holds with high probability:

$$\mathcal{R}^{\alpha,\beta}_{AUC}(f_{\boldsymbol{\theta}}, \mathcal{S}) \leq \hat{\mathcal{R}}^{\ell}_{\psi}(f_{\boldsymbol{\theta}}, \mathcal{S}) + \tilde{\mathcal{O}}\left(\left(\frac{\text{VC}}{n_+}\right)^{1/2} + \left(\frac{\text{VC}}{n_-}\right)^{1/2}\right)$$

where $\tilde{\mathcal{O}}$ *is the big-O complexity notation hiding the logarithm factors,*

$$\mathcal{R}^{\alpha,\beta}_{AUC}(f_{\boldsymbol{\theta}}, \mathcal{S}) = 1 - \text{AUC}^{\beta}_{\alpha}(f_{\boldsymbol{\theta}}, \mathcal{S}),$$

and VC *is the VC dimension of the hypothesis class:*

$$\mathcal{T}(\mathcal{F}) \triangleq \{\text{sign}(f_{\boldsymbol{\theta}}(\cdot) - \delta) : f_{\boldsymbol{\theta}} \in \mathcal{F}, \delta \in \mathbb{R}\}$$

- Two instantiations

Example 1 (Polynomial Surrogate Model).

$$\varphi_\gamma^{\text{poly}}(t) = \frac{1}{\gamma} \cdot t^\gamma, \psi_\gamma^{\text{poly}}(t) = t^{\frac{1}{\gamma-1}}, \gamma > 2$$

Example 2 (Exponential Surrogate Model).

$$\varphi_\gamma^{\exp}(t) = \frac{(1-t)(\log(1-t)-1)+1}{\gamma}$$

$$\psi_\gamma^{\exp}(t) = 1 - e^{-\gamma t}$$

Experiments

Table 1. Performance Comparisons over different metrics and datasets, where (x, y) stands for TPAUC(x, y) in short.

dataset	type	methods	Subset1 (0.3,0.3)	Subset1 (0.4,0.4)	Subset1 (0.5,0.5)	Subset2 (0.3,0.3)	Subset2 (0.4,0.4)	Subset2 (0.5,0.5)	Subset3 (0.3,0.3)	Subset3 (0.4,0.4)	Subset3 (0.5,0.5)
CIFAR-10-LT	Competitors	CE-RW	9.09	30.86	47.99	72.83	83.33	88.71	23.47	44.44	59.69
		Focal	9.84	30.89	50.83	75.72	85.10	90.06	21.47	45.88	59.09
		CBCE	3.29	27.30	43.95	69.48	80.80	86.87	12.94	34.06	51.09
		CBFocal	9.04	31.73	48.13	77.99	86.75	91.13	21.32	43.03	59.11
		SqAUC	18.05	40.74	57.94	80.09	87.78	91.87	31.52	50.00	64.42
	Ours	Poly	**21.43**	**44.41**	**59.10**	80.66	**88.07**	**92.15**	**36.54**	**54.48**	67.19
		Exp	20.86	41.78	58.38	**81.22**	87.88	91.93	32.47	53.86	**67.32**
CIFAR-100-LT	Competitors	CE-RW	31.43	52.60	66.21	79.70	88.06	92.64	3.09	21.32	40.75
		Focal	36.51	61.71	73.25	83.08	90.35	93.76	8.09	28.88	49.89
		CBCE	17.53	38.79	55.19	67.91	79.32	85.82	1.84	18.46	37.04
		CBFocal	41.85	62.41	73.13	82.75	89.57	92.89	7.10	29.12	44.84
		SqAUC	63.24	76.62	84.68	91.02	93.69	94.73	41.60	60.36	70.86
	Ours-TPAUC	Poly	**68.02**	**79.11**	**85.17**	**91.13**	**93.78**	**95.69**	**47.07**	**65.89**	**75.08**
		Exp	63.24	77.94	84.62	90.69	93.74	95.41	44.54	64.58	73.02
Tiny-ImageNet-200-LT	Competitors	CE-RW	80.90	87.76	91.54	93.30	96.15	97.53	90.37	94.34	96.75
		Focal	81.18	88.06	91.72	93.23	96.08	97.59	91.35	94.87	96.63
		CBCE	80.64	87.58	91.17	93.77	**96.52**	**97.77**	91.66	95.19	96.79
		CBFocal	80.44	87.95	91.91	93.46	96.43	97.64	91.06	94.82	96.62
		SqAUC	80.16	87.99	91.67	93.10	96.07	97.32	**92.15**	95.16	96.75
	Ours-TPAUC	Poly	80.44	88.21	91.98	93.00	95.61	97.47	92.02	**95.25**	**96.84**
		Exp	**82.61**	**89.13**	**92.62**	**93.82**	96.12	97.38	91.25	94.78	96.57

Personal homepage: https://joshuaas.github.io (with QR code)

Email: yangzhiyong@iie.ac.cn

Group homepage: https://qmhuang-ucas.github.io