Understanding the Ranking Loss for Recommendation with Sparse User Feedback

Zhutian Lin[1⋆], **Junwei Pan**[2⋆], Shangyu Zhang[2], Ximei Wang[2], Xi Xiao[1], Shudong Huang[2], Lei Xiao[2], Jie Jiang[2],

KDD2024 BARCELONA, SPAIN

CTR Prediction: Binary Classification

- **Binary Cross Entropy (BCE) loss** for CTR prediction

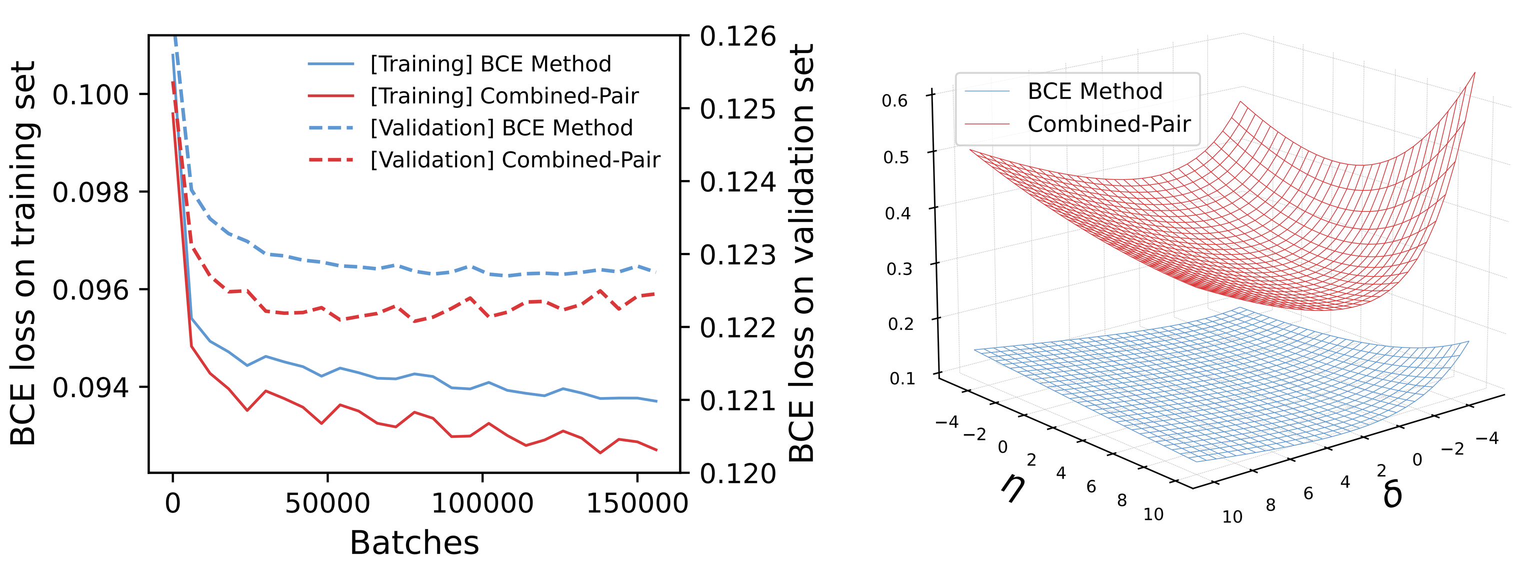
$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\sigma(z_i)) + (1-y_i)\log(1-\sigma(z_i))], \tag{1}$$

- **Classification-ranking combination loss**: combining the BCE loss with a ranking loss

$$\mathcal{L}^{\mathrm{CP}} = \alpha\mathcal{L}_{\mathrm{BCE}} + (1-\alpha)\mathcal{L}_{\mathrm{RankNet}}, \tag{2}$$

$$\mathcal{L}_{\mathrm{RankNet}} = -\frac{1}{N_+N_-}\sum_{i=1}^{N_+}\sum_{j=1}^{N_-}\log(\sigma(z_i^{(+)} - z_j^{(-)})). \tag{3}$$

Investigation of Classification Ability



Finding 1. Combined-Pair gets a lower BCE loss than the BCE method on the validation set, indicating that it improves the classification ability rather than only the ranking ability.

Finding 2. Combined-Pair gets a lower BCE loss than the BCE method on the training set, indicating that involving an auxiliary ranking loss helps the optimization of the BCE loss.

Gradient Analysis

- **Gradient of BCE loss for negative sample**

$$\begin{aligned}\nabla_{z_j^{(-)}}\mathcal{L}_{\mathrm{BCE}} &= \frac{1}{1-\sigma(z_j^{(-)})}\cdot\sigma(z_j^{(-)})(1-\sigma(z_j^{(-)})) \\ &= \sigma(z_j^{(-)}) = \hat{p}_j.\end{aligned} \tag{4}$$

Finding 3. When positive feedback is sparse, the gradients of negative samples vanish since they are proportional to the estimated positive rates, which are small in an unbiased estimator.

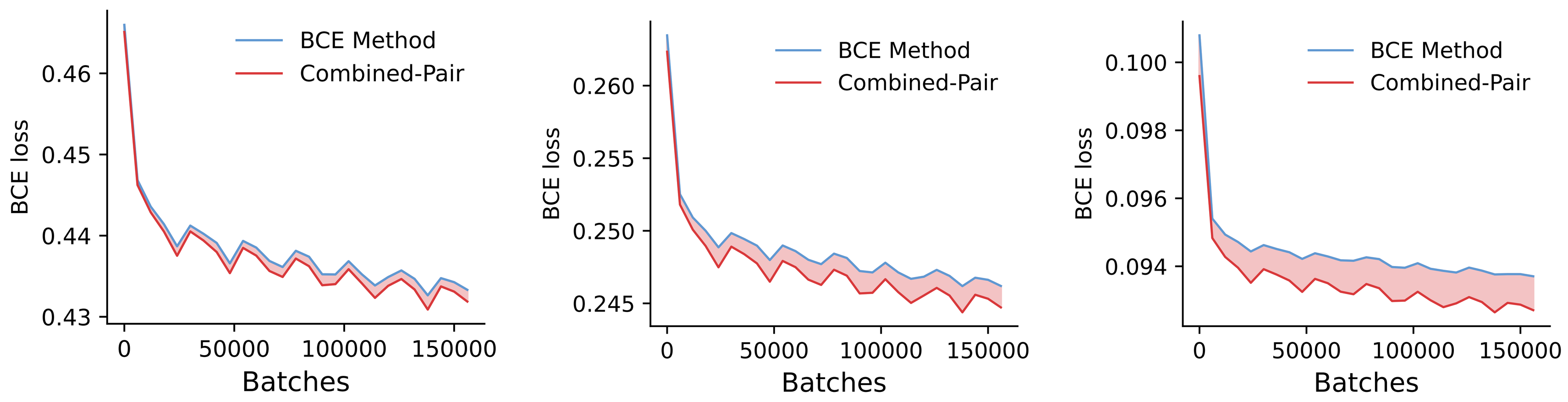
- **Gradient of RankNet loss for negative sample**

$$\begin{aligned}\nabla_{z_j^{(-)}}\mathcal{L}^{\mathrm{CP}}_{\mathrm{Rank}} &= \frac{1}{N_+}\sum_{i=1}^{N_+}\sigma(z_j^{(-)} - z_i^{(+)}) \\ &> \sigma(z_j^{(-)}) = \nabla_{z_j^{(-)}}\mathcal{L}_{\mathrm{BCE}},\end{aligned} \tag{5}$$

Finding 4. When positive feedback is sparse, Combined-Pair has larger gradients for negative samples than the BCE method.

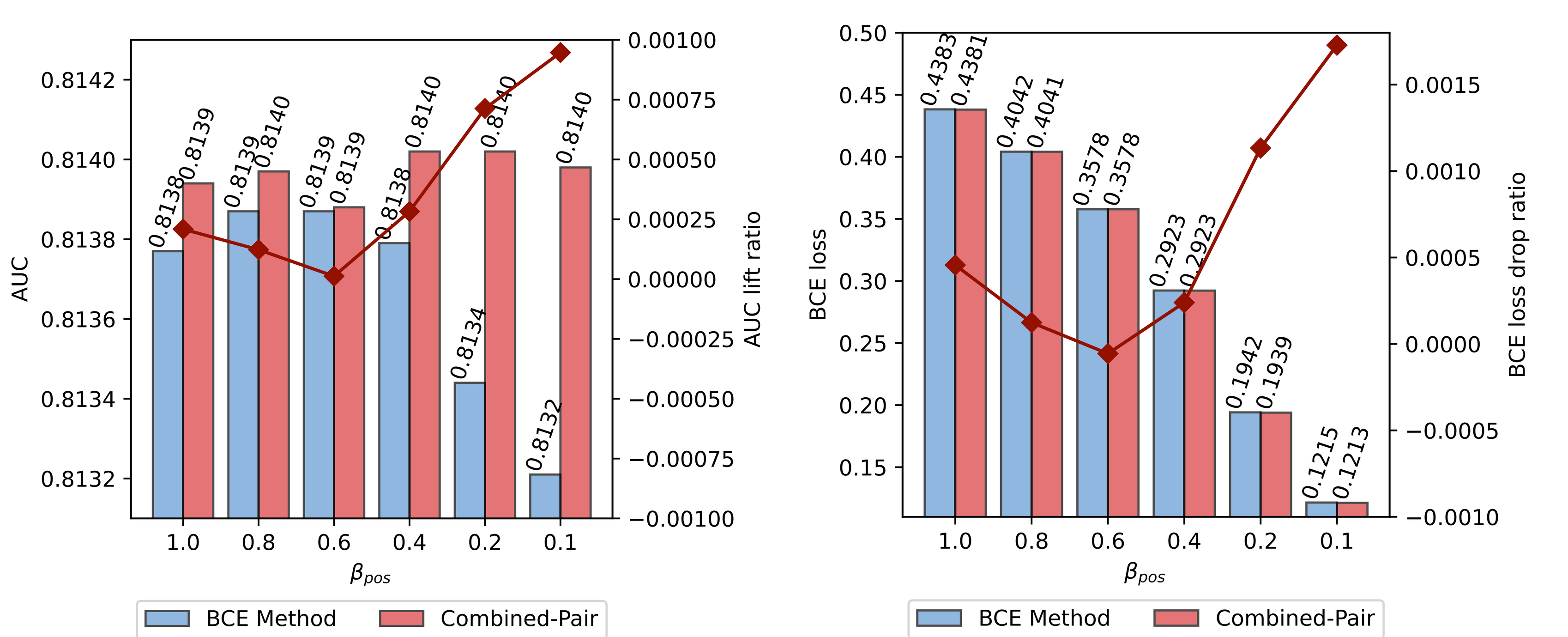
Performance Evaluation with various Positive Sparsity Rates

- **Training BCE loss under varying positive sparsity rates.**



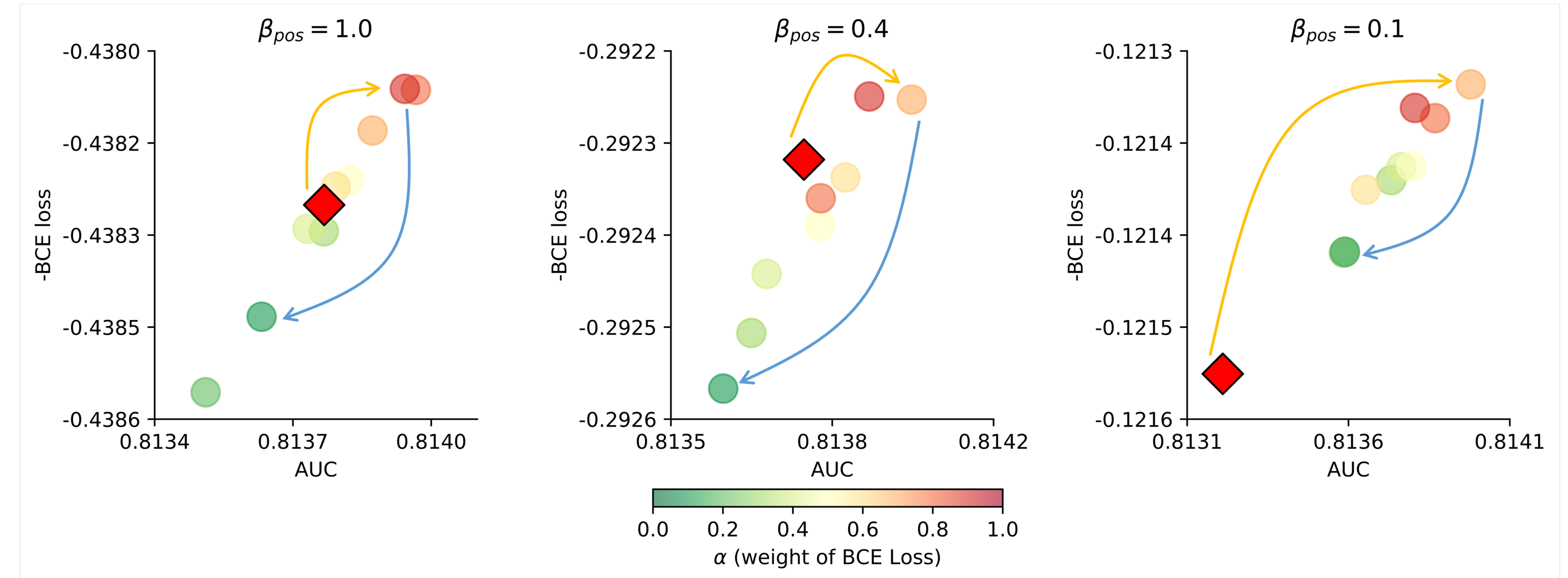
(a) $\beta_{\mathrm{pos}} = 1.0$ (b) $\beta_{\mathrm{pos}} = 0.4$ (c) $\beta_{\mathrm{pos}} = 0.1$

- **Performance evaluation under varying positive sparsity rates.**



Trade-off between Classification Loss and Ranking Loss

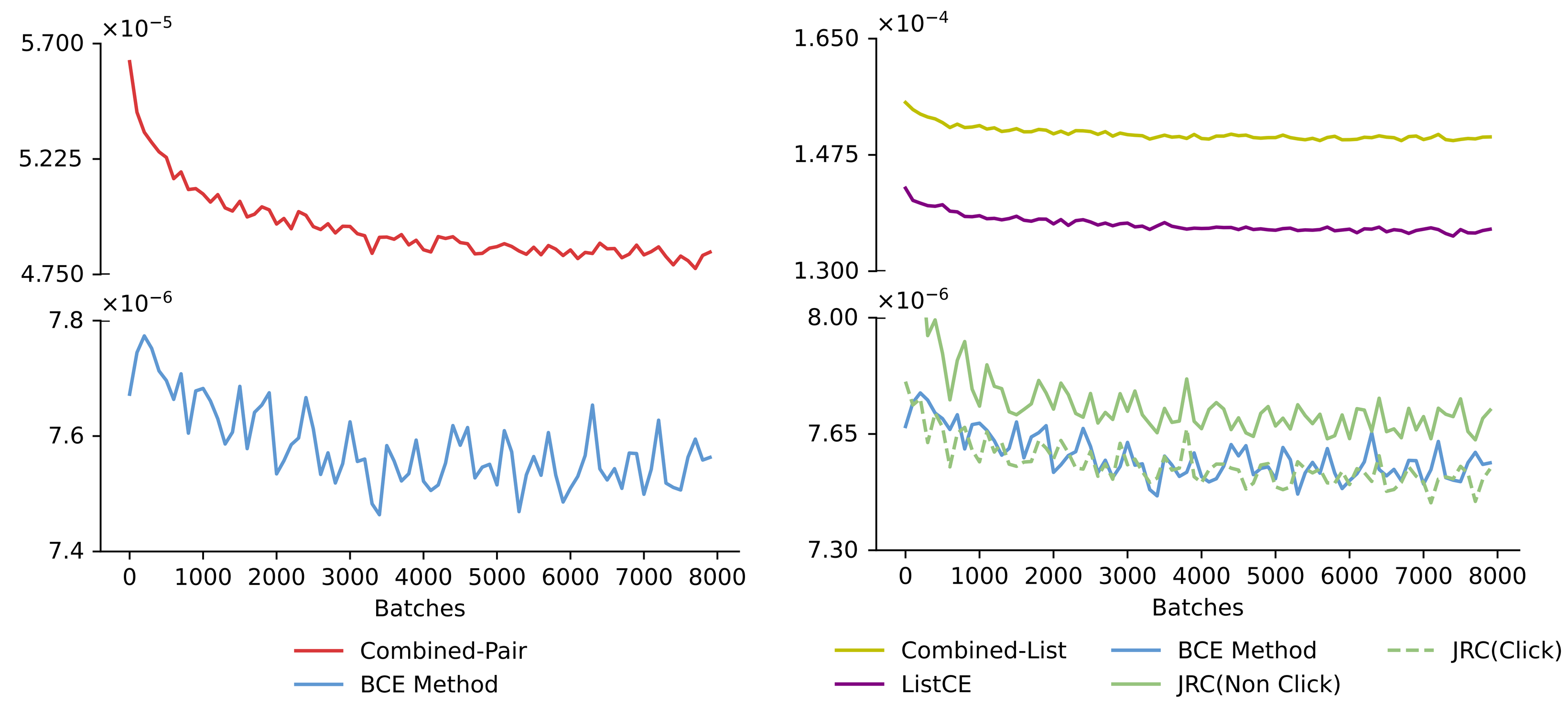
- **AUC and negative BCE Loss of Combined-Pair and BCE method with various α**



- The classification and ranking abilities can be improved monotonically by decreasing α up to a certain threshold.
- When α is further decreased. In other words, as the ranking loss becomes more dominant in the combination loss beyond a certain threshold, both the classification and ranking abilities deteriorate monotonically.

Evaluation of Different Ranking Losses

- **Gradient Norm of Various Ranking Losses**



- **Performance Evaluation of Various Ranking Losses**

Metric	BCE	BCE+Pairwise	BCE+Listwise		
		Combined-Pair	JRC	Combined-List	RCR
AUC↑	0.81321	**0.81398**↑	0.81355↑	0.81351↑	0.81349↑
BCE Loss↓	0.12152	**0.12131**↓	0.12146↓	0.12152	0.12141↓

Beyond Ranking Loss

- **Focal Loss**

$$\mathcal{L}_{\mathrm{Focal}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i(1-\hat{p}_i)^{\gamma}\log(\hat{p}_i) + (1-y_i)\hat{p}_i^{\gamma}\log(1-\hat{p}_i))]$$

- **Combined Contrastive**

$$\mathcal{L}^{\mathrm{CC}} = \alpha\mathcal{L}_{\mathrm{BCE}} + (1-\alpha)\mathcal{L}_{\mathrm{Contr}}, \tag{6}$$

$$\mathcal{L}_{\mathrm{Contr}} = \frac{1}{|N|}\sum_{i=1}^{N}\frac{-1}{|P(i)|}\sum_{p\in P(i)}\log\frac{\exp(\mathbf{z}_i\mathbf{z}_p/\tau)}{\sum_{a\in A(i)}\exp(\mathbf{z}_i\mathbf{z}_a/\tau)} \tag{7}$$

Stage	Metrics	BCE Method	Combined-Contrastive
Training	Gradient Norm	4.9×10^{-6}	$\mathbf{7.5\times10^{-6}}$
	BCE Loss ↓	0.09667	**0.09428**↓
Testing	AUC↑	0.81321	**0.81340**↑
	BCE Loss ↓	0.12152	**0.12147**↓

Online Experiments

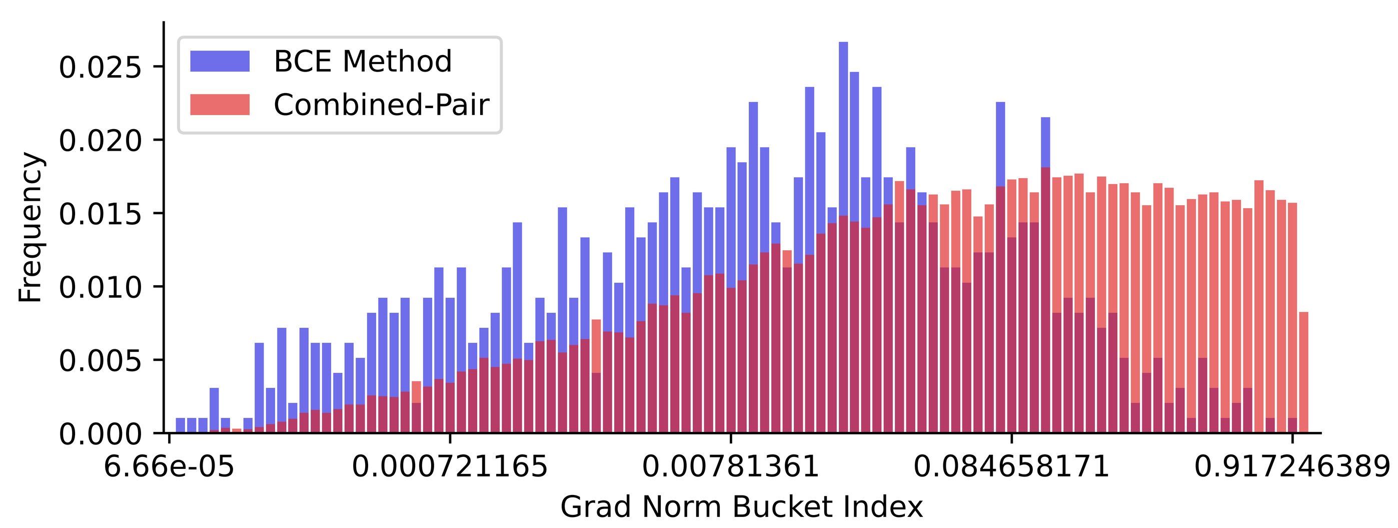
Ad Scenario	CTR	GMV	Cost
WeChat Channels	+0.91%	+1.08%	+0.29%
WeChat Moment	+0.16%	+0.70%	+0.59%
DSP	-0.04%	+0.55%	+0.15%

Table: Online A/B Testing Results.

Launch Date	GMV	Cost
T	+1.04%	+0.27%
T-1	+1.04%	+0.27%
T-2	+0.83%	+0.47%
T-3	+0.81%	+0.17%
Total	+1.26%	+0.34%

Table: Online A/B Testing Results for New Ads.

- **Distribution of gradient norms for negative samples in online experiments**



Tsinghua University, Tencent Inc.

Code: `https://github.com/SkylerLinn/Understanding-the-Ranking-Loss`

Mail: jonaspan@tencent.com