

1 RESPONSE TO REVIEWER PBE8

Dear Reviewer PBe8, we sincerely thank you for your valuable feedback on our submission. Below is our responses to the concerns you raised. We have incorporated the following contents into the updated version of our paper, which we believe will help enhance the quality of our submission.

Q1: The paper can be improved with more performance comparison with existing work and state-of-the-art performance.

A1: We agree with your suggestion that comparing UNICBE with state-of-the-art methods can help fully validate its effectiveness. In our experiments, we select the most widely used state-of-the-art methods, Arena and AlpacaEval, for comparison to evaluate the performance of UNICBE. In fact, although comparing-based evaluation holds significant research value in the era of large language models, the existing research in this area remains limited. Arena and AlpacaEval represent the most relevant and practical works we could identify. We encourage future researchers to explore this field further and investigate more efficient ways to utilize valuable preference signals.

Q2: Expect more statistical and experimental conclusions with the proposed CBE method for the scenario of large-scale preference learning.

A2: We agree with your suggestion and therefore attempt to conduct experiments in a large-scale, highly dynamic CBE scenario that is closer to real-world conditions. Specifically: Starting with a sample size of $N = 600$ and model number of $M = 12$, we execute a random operation at each time step. The operations included: adding one model to be evaluated with a probability of 0.01, removing one model with a probability of 0.01, adding one potential sample with a probability of 0.01, randomly deleting one sample with a probability of 0.01, and taking no action with a probability of 0.96. Based on the experimental results shown in Figure 1, we have the following observations:

- The convergence speed of all baseline methods significantly slowed down. None of the baseline methods achieve a Spearman correlation coefficient of 0.96 or a Pearson correlation coefficient of 0.97 by $T = 2000$, highlighting the difficulty of model evaluation in this setting. In contrast, UNICBE achieve rapid convergence, reaching a Spearman coefficient of approximately 0.97 and a Pearson coefficient exceeding 0.98 by $T = 2000$.
- Over the long term, as T increases, UNICBE consistently demonstrates over 10% savings in preference budget across all metrics, even under this challenging setting, showcasing its strong practicality.
- An interesting observation is that ALPACAEVAL exhibits better convergence in the early stages compared to RANDOM and ARENA, supporting our previous conclusions in Table 1 (original manuscript). However, as T increases, ALPACAEVAL’s lack of accuracy optimization objective leads to its performance being surpassed by RANDOM and ARENA.

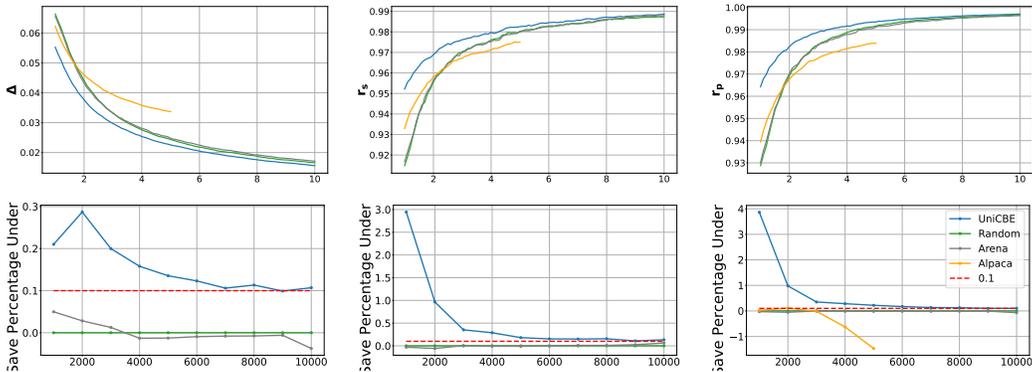


Figure 1: Results of compared CBE methods in a scenario where models and samples are dynamically added or removed at a random frequency.