

Relative Advantage Debiasing for Watch-Time Prediction in Short-Video Recommendation

Emily Liu, Kuan Han*, Minfeng Zhan, Bocheng Zhao, Guanyu Mu, Yang Song

ByteDance Inc.

{emily.zfliu, kuan.han, zhanminfeng, zhaobocheng, muzhuoyu}@bytedance.com, ys@sonyis.me

Abstract

Watch time is widely used as a proxy for user satisfaction in video recommendation platforms. However, raw watch times are influenced by confounding factors such as video duration, popularity, and individual user behaviors, potentially distorting preference signals and resulting in biased recommendation models. We propose a novel relative advantage debiasing framework that corrects watch time by comparing it to empirically derived reference distributions conditioned on user and item groups. This approach yields a quantile-based preference signal and introduces a two-stage architecture that explicitly separates distribution estimation from preference learning. Additionally, we present distributional embeddings to efficiently parameterize watch-time quantiles without requiring online sampling or storage of historical data. Both offline and online experiments demonstrate significant improvements in recommendation accuracy and robustness compared to existing baseline methods.

Extended version — <https://arxiv.org/abs/2508.11086>

Introduction

Immersive video viewing experience, such as TikTok and Reels, allows users to dive into a continuous stream of content that captures attention through full-screen visuals and intuitive swipe-based interactions. It also brings unique challenges to recommender systems compared to traditional ones, as explicit interaction signals are often missing (e.g., clicks, ratings) or sparse (e.g., likes, comments, shares) (Lin et al. 2023; Davidson et al. 2010). This scarcity reduces the effectiveness of explicit feedback for learning user preferences. As a result, the duration a user spends viewing a video, commonly referred to as watch time, has become the standard implicit proxy for user interest. While watch time provides a continuous and behaviorally grounded measure of engagement, it is inherently confounded by factors unrelated to genuine preference (Zhan et al. 2022). For example, longer videos accumulate higher watch times regardless of true interest, leading to duration bias and a tendency to over-recommend long-form content (Zhan et al. 2022; Zhao et al. 2024). Such biases can distort preference

estimation and undermine recommendation quality (Zheng et al. 2022). To overcome these limitations, robust debiasing strategies are needed to recover authentic user interests and ensure both fairness and effectiveness in recommendations (Wang et al. 2021b).

Prior work has primarily targeted duration bias by partitioning watch-time into length-based buckets and applying transformations such as quantile normalization (Zhan et al. 2022), residual-gain adjustment (Zheng et al. 2022; Tang et al. 2023), and counterfactual or nonlinear mappings (Zhao et al. 2024). While these techniques effectively mitigate video-length distortions, they offer limited protection against other confounders such as content popularity or variations in user engagement patterns. More recent approaches model the full conditional watch-time distribution by conditioning on each user–video pair to address variability in watch-time estimation (Lin et al. 2025a), and may further incorporate additional context such as demographics and content category (Lin et al. 2025b). While this comprehensive approach captures uncertainty and heterogeneity, it suffers from the “single-observation” problem: in practice, users rarely replay the same video, so there is almost never more than one sample under identical conditions. Consequently, distribution estimates derived from these methods tend to be noisy, prone to overfitting, and may introduce unintended dependencies between distribution fitting and recommendation goals, ultimately undermining preference modeling.

To address these challenges, we propose a relative advantage framework that debiases watch time by mapping each observed watch time onto two empirical reference distributions: one conditioned on video ID (aggregating all viewers’ watch times for that video) and the other on user ID (aggregating that user’s watch times across videos). Converting a watch time into its empirical quantile within each umbrella-conditioned distribution yields a uniform, bias-corrected signal that reflects relative engagement within each context. These umbrella factors can either be applied individually to correct video- and user-level confounders such as video length, popularity, and individual viewing habits, or combined through Bayesian evidence fusion to produce a robust and calibrated preference score.

*Correspondence to kuan.han@bytedance.com
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

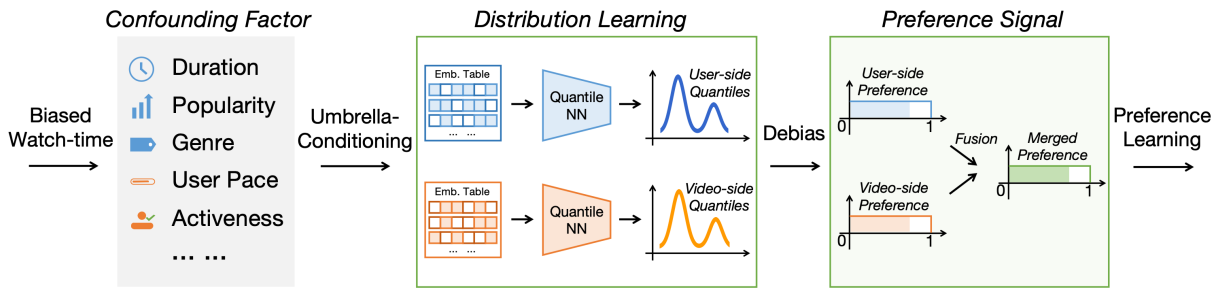


Figure 1: An overview of the RAD framework. First, distributional embeddings for user- and video-side confounders are learned from historical data. Then, these embeddings are used to measure the CDF of a biased watch-time value conditioned on user and video confounding variables. User-side and video-side signals are merged through Bayesian fusion to create a single bias-free label for preference learning.

This design consists of two stage. The first stage involves estimating conditioned distributions; the second stage involves modeling preferences, ensuring modularity, numerical stability, and interpretability. Both offline benchmarks and live A/B tests demonstrate significant improvements in recommendation accuracy and robustness compared to existing baselines.

The major contributions of this work are:

- A novel relative advantage debiasing framework for implicit watch-time feedback, correcting both item- and user-level confounders beyond simple duration bias.
- A two-phase architecture that explicitly decouples distribution estimation from preference modeling, enhancing training stability and model interpretability.
- A distributional embedding method to parameterize watch-time quantiles directly, eliminating the need for additional online indexing or historical data storage.
- Comprehensive offline and online evaluations demonstrating significant improvements in recommendation accuracy, robustness, and fairness over state-of-the-art baselines.

Related Work

Watch-time Prediction Predicting video watch time is fundamental for video recommendation systems, as it serves as a primary proxy for user engagement. Early models, such as Weighted Logistic Regression (WLR) (Covington, Adams, and Sargin 2016) weight impressions by view duration but fail to handle the heavy-tailed and varying video length distributions, which are typical of short-form content. To better capture ordinal and uncertain aspects of watch time, Tree-based Progressive Regression (TPM) (Lin et al. 2023) discretizes watch time into ordered intervals and uses a hierarchical tree of binary classifiers. By contrast, CREAD (Sun et al. 2024) introduces error-adaptive bucketization to balance classification and restoration, effectively handling the long-tailed distributions commonly observed in streaming platforms.

Watch-time Debiasing A range of duration-debiasing methods have recently emerged. DVR (Zheng et al. 2022) introduces Watch-Time-Gain, normalizing watch time within duration groups using adversarial learning. D2Q (Zhan et al. 2022) applies causal backdoor adjustments to control for duration effects, while CVRDD (Tang et al. 2023) uses counterfactual inference to directly eliminate duration bias at prediction time. Further developments also address measurement noise: D2Co (Zhao et al. 2023) simultaneously corrects for duration bias and noisy watch-time through GMM-based bias/noise estimation and correction, and CWM (Zhao et al. 2024) introduces counterfactual watch time to recover engagement truncated by video duration. In contrast to these approaches, our method integrates both item-side and user-side relative advantage signals to jointly correct a range of biases, such as duration, popularity, and engagement patterns, within a single unified model, paving the way for a more generalizable and scalable debiasing framework.

Quantile Regression Quantile regression estimates conditional quantiles (such as the median or tail probabilities), providing a more comprehensive characterization of outcome distributions than summary statistics such as the conditional mean, especially in non-Gaussian or heavy-tailed data (Koenker and Bassett 1978; Meinshausen 2006; Angrist, Chernozhukov, and Fernandez-Val 2004). This technique has been effectively incorporated into machine learning models, including neural networks (Padilla, Tansey, and Chen 2022) and random forests (Meinshausen 2006), to yield uncertainty estimates and enhance robustness. In the context of video recommendation, D2Q (Zhan et al. 2022) was among the first to leverage quantile regression within duration buckets to correct for duration bias in watch-time prediction. Conditional Quantile Estimation (CQE) (Lin et al. 2025a) extends this approach by predicting multiple watch-time quantiles per user-video tuple, enabling more flexible and context-aware modeling. AlignPxtr (Lin et al. 2025b) further extends this paradigm by aligning predicted quantiles across various bias conditions (such as duration, demographics, or content category) through quantile mapping, further separating user interest from confounding ef-

fects. However, both CQE and AlignPxr estimate conditional quantiles from only a single watch-time observation per user–video pair, which can lead to unstable and overfitted quantile estimates. In contrast, our approach aggregates watch-time distributions on both the item and user sides, enabling more robust and generalizable relative-advantage metrics. This separation of distribution estimation from preference modeling allows us to capture uncertainty and correct for multiple biases within a unified framework.

Method

A Generative Model of Watch Time

We treat each observed watch time $S_{u,i}$ as a generative model with two sources: the true underlying preference user u has for video i , and any systematic confounding factors outside of this preference, formally as

$$S_{u,i} = C_{u,i} + P_{u,i} + \varepsilon_{u,i}, \quad (1)$$

where $C_{u,i}$ captures the aggregate effect of confounders, $P_{u,i}$ reflects the genuine user–video preference, and $\varepsilon_{u,i}$ denotes irreducible noise.

To make the impact of confounding explicit, let \mathcal{C} represent the collection (or σ -algebra) generated by all fine-grained confounders. Applying the law of total variance, the variability in observed watch time decomposes as

$$\text{Var}(S_{u,i}) = \underbrace{\text{Var}(\mathbb{E}[S_{u,i} | \mathcal{C}])}_{\text{bias variance}} + \underbrace{\mathbb{E}[\text{Var}(S_{u,i} | \mathcal{C})]}_{\text{residual variance}}. \quad (2)$$

Here, $\text{Var}(\mathbb{E}[S_{u,i} | \mathcal{C}])$ quantifies the bias variance, i.e., systematic differences in watch time induced by confounders, while $\mathbb{E}[\text{Var}(S_{u,i} | \mathcal{C})]$ is the residual variance attributable to true preferences or irreducible noise. By conditioning on confounders, we seek to minimize the bias variance, enabling the model to focus on the residual variation that more accurately reflects genuine user–video preferences and results in more stable, interpretable preference learning.

Umbrella Conditioning for Variance Reduction

In large-scale recommendation systems, the sheer number of confounders makes it impractical to correct for each one individually. However, many real-world confounders are deterministic functions of higher-level identifiers—what we refer to as *umbrella factors*. For instance, every video-side confounder—such as duration, category, creator, or popularity—denoted $c^{(1)}, \dots, c^{(K)}$, is determined by the video ID i . For K confounders dependent on video ID, let $\sigma(c^{(1)}, \dots, c^{(K)})$ be the σ -algebra generated by these fine-grained confounders, and $\sigma(i)$ that generated by video ID. Since

$$\sigma(c^{(1)}, \dots, c^{(K)}) \subseteq \sigma(i), \quad (3)$$

we have the following variance reduction property.

Proposition 1 (Variance monotonicity under umbrella sufficiency). *For any video-side confounder $c^{(k)}$,*

$$\text{Var}(\mathbb{E}[S_{u,i} | i]) \geq \text{Var}(\mathbb{E}[S_{u,i} | c^{(k)}]). \quad (4)$$

Proof of this statement is provided in the Appendix. Intuitively, conditioning on video ID collapses confounders—such as duration, category, creator, and popularity—into a single context; the same logic applies to user ID, which subsumes factors like activeness level, device, and viewing habits. While any conditioning variable reduces bias variance to some extent, the amount removed depends on how much between-group variation it captures. Since user and video IDs encompass a broad range of real-world confounders, umbrella conditioning offers a scalable and effective way to produce cleaner residual signals for robust preference modeling.

Conditional Quantile Transformation

To achieve robust debiasing, we transform each observed watch time by its conditional cumulative distribution function (CDF), conditioned on an umbrella factor such as video or user ID. Let $F_{S|G}(s) = \Pr(S \leq s | G)$ denote the conditional CDF for umbrella factor $G \in \{i, u\}$. The within-context quantile is then defined as

$$Q_{u,i}(G) = F_{S|G}(S_{u,i}) \in (0, 1). \quad (5)$$

Proposition 2 (Statistics and Independence of Quantile Labels). *For any umbrella factor G , the quantile label $Q_{u,i}(G)$ has mean $\mathbb{E}[Q_{u,i}(G)] = \frac{1}{2}$ and variance $\text{Var}(Q_{u,i}(G)) = \frac{1}{12}$, and is statistically independent of G , i.e., $Q_{u,i}(G) \perp\!\!\!\perp G$.*

A proof of Proposition 2 is provided in the Appendix. This mapping ensures that quantile labels are bias-free, bounded, and homoskedastic, stabilizing both the quantile signal and optimization process. Notably, after this transformation, $Q(G)$ is statistically independent of confounders included in G —that is, all systematic group-level biases are removed, and the remaining variation reflects latent user–video preference. These properties make quantile (CDF) labels an effective and principled choice for preference modeling over alternatives like z-scores.

User-Side Duration Heterogeneity Unlike video ID, which encapsulates duration as an intrinsic attribute, user ID cannot fully account for duration—an especially dominant source of bias in watch-time (Zhan et al. 2022). To address this, we first split all watch times into D near-equal-mass bins (e.g., $D = 4$) based on duration (Zhan et al. 2022). For each bin, we estimate the empirical user-side CDF $F_{S|u,\text{bin}}(s)$, and assign the bin-specific quantile label to each watch time:

$$Q_{u,i}^{(\text{user},\text{bin})} = F_{S|u,\text{bin}}(S_{u,i}) \in (0, 1). \quad (6)$$

This additional conditioning ensures that user-side RAD labels remain robust to heterogeneity in video length—a correction not required for video ID umbrella as duration is already captured.

Relative Advantage Debiasing (RAD)

Based on the conditional quantile transformation introduced above, we propose **Relative Advantage Debiasing (RAD)**, a two-stage procedure leveraging hierarchical umbrella conditioning. RAD transforms watch times into quantile-based labels, providing stable, bounded, and unbiased signals of relative user engagement.

Stage 1: RAD Label Estimation. For each umbrella factor (video or user), we estimate empirical watch-time distributions from historical logs and derive quantile labels:

- **Video-side RAD (RAD-V):** For each video i , we compute its empirical CDF, $F_{S|i}(s)$, as quantile labels:

$$Q_{u,i}^{(\text{video})} = F_{S|i}(S_{u,i}) = \Pr_{u'}(S_{u',i} \leq S_{u,i} | i). \quad (7)$$

- **User-side RAD (RAD-U):** We partition watch times into D bins, and compute the empirical CDF specific to the d -th bin as quantile labels:

$$Q_{u,i}^{(\text{user})} = F_{S|u,d}(S_{u,i}) = \Pr_{i'}(S_{u,i'} \leq S_{u,i} | u, d). \quad (8)$$

Stage 2: Preference Modeling. Given these debiased RAD labels, we train a parametric model (e.g., MLP, DCN, etc) to predict RAD-U or RAD-V label. Separating two stages yields disentangled, stable targets to learn.

We evaluate RAD with two hypotheses. First, the bounded and homoskedastic nature of RAD labels makes training more stable and efficient, leading to lower watch-time prediction error (e.g., reduced MAE) when mapped back to raw values. We validate this by comparing RAD-based models with standard baselines on watch-time prediction. Second, by removing systematic biases, RAD labels offer a cleaner measure of user engagement, which should yield better ranking performance—higher XAUC/XGAUC and improved online metrics. We confirm this by ranking with RAD predictions and measuring gains in both offline and online evaluation.

Dual-sided Bayesian Evidence Fusion

While RAD-U and RAD-V each correct user- and video-side biases, fusing them yields a unified, robust preference score. We achieve this by first mapping each quantile label into z-score space using the probit (inverse normal CDF) transform:

$$z_{u,i}^{(\text{user})} = \Phi^{-1}(Q_{u,i}^{(\text{user})}), \quad z_{u,i}^{(\text{video})} = \Phi^{-1}(Q_{u,i}^{(\text{video})}). \quad (9)$$

We then combine the two z-scores through a weighted average, and mapping the fused score back to quantile space::

$$\hat{z}_{u,i} = \frac{\alpha z_{u,i}^{(\text{user})} + \beta z_{u,i}^{(\text{video})}}{\sqrt{\alpha^2 + \beta^2}} \rightarrow Q_{u,i} = \Phi(\hat{z}_{u,i}), \quad (10)$$

where α and β are precision weights that reflect the reliability of each signal. In offline experiments, we set these weights proportional to the statistical support, that is, the number of samples used to estimate the user- and video-side CDFs. In large-scale or online deployment, we use equal weights ($\alpha = \beta = 1$), since both user and video CDFs are well-supported with sufficient data and this simplifies deployment.

This Bayesian fusion produces a stable, calibrated, and interpretable score that jointly corrects user- and video-side biases, naturally weighting each view by its confidence. The unified label integrates directly into downstream models and improves robustness, especially in cold-start scenarios where one side has limited data. The effect of the ratio β/α on results, and the robustness of RAD to distributional shift, is discussed in the Appendix.

Experiments

We show that Relative Advantage Debiasing (RAD) significantly improves both watch-time prediction accuracy and ranking quality over state-of-the-art baselines. Further analysis confirms the effectiveness of RAD’s two-stage learning and Bayesian fusion, and highlights an efficient strategy for distribution learning. Finally, large-scale online A/B tests demonstrate clear gains in user engagement, underscoring RAD’s real-world impact.

Dataset and Experiment Setup

Datasets We conduct offline experiments on two large-scale short-video datasets: **(1) KuaiRand-Pure:** A public benchmark from the KuaiShou platform, widely used for debiasing tasks in sequential video recommendation. We use the KuaiRand-Pure subset as recommended by prior work (Gao et al. 2022), chronologically splitting data using a sliding timestamp cut-off, allocating training (79.6%), validation (8.7%), and test (11.6%) sets. Users in validation or test sets are filtered to ensure they also appear in training. The final split contains 26,592 users, 7,146 items, and 1,384,425 user–video interactions. **(2) An Offline Industrial Dataset:** We further evaluate RAD on a offline industrial dataset sampled from a large-scale short-video platform. The snapshot contains well over a billion user–video interactions, making it orders of magnitude larger than KuaiRand. Data are split chronologically: the first 14 days form the training set, and the following day is used for testing. This large-scale setting allows us to assess whether RAD’s improvements persist under high-data conditions.

Baselines We compare our method against a comprehensive set of state-of-the-art baselines covering direct regression, normalization, noise correction, and quantile-based modeling. Baselines include: Value Regression (VR), Play Completion Rate (PCR) (Zhao et al. 2024), Weighted Logistic Regression (WLR) (Covington, Adams, and Sargin 2016), Watch-time Gain (WTG) (Zheng et al. 2022), Debiased and Denoised Correction (D2Co)(Zhao et al. 2023), Duration-debiased Quantiles (D2Q) (Zhan et al. 2022), Counterfactual Watch Model (CWM) (Zhao et al. 2024), and Conditional Quantile Estimation (CQE) (Lin et al. 2025a). Detailed descriptions of baselines are given in the appendix, while implementations follow the corresponding papers and established protocols (Zhao et al. 2024; Lin et al. 2025a).

Backbones and Hyperparameters All methods are implemented with the following backbone architectures to test generalizability, including multi-layer perceptron (MLP), the state-of-the-art Deep & Cross Network (DCN) (Wang et al. 2017), and its recent extensions, DCN_v2 (Wang et al. 2021a) and GDCN (Wang et al. 2023). A description of hyperparameters is given in the Appendix.

Evaluation Metrics We evaluate models using standard metrics from literature (Zhan et al. 2022), including: (1) mean absolute error (MAE), which measures the average absolute difference between predicted and true watch times; (2) XAUC, an extension of AUC for continuous outcomes to assess how well model predictions preserve the ranking

Metric	Backbone	VR	PCR	WLR	WTG	D2Co	D2Q	CWM	CQE	RADV	RADU	RADUV
MAE	MLP	21.525	45.912	21.229	22.627	22.123	19.763	20.330	21.434	18.050	18.221	18.050
	DCN	23.332	46.095	21.369	23.225	23.046	19.888	19.906	21.672	18.114	18.223	18.088
	DCNV2	23.120	45.864	21.376	22.718	21.990	19.823	20.026	21.245	18.083	18.213	18.068
	GDCN	23.718	45.899	21.388	22.820	22.235	19.808	20.693	21.600	18.071	18.218	18.058
XAUC	MLP	0.6781	0.5679	0.4566	0.6925	0.6978	0.6888	0.7096	0.7000	0.7137	0.7151	0.7178
	DCN	0.6871	0.5672	0.3771	0.6880	0.6900	0.6863	0.7099	0.7037	0.7134	0.7160	0.7181
	DCNV2	0.6845	0.5673	0.3856	0.6922	0.6978	0.6864	0.7092	0.7034	0.7119	0.7150	0.7172
	GDCN	0.6712	0.5673	0.3907	0.6911	0.6964	0.6869	0.7091	0.7015	0.7140	0.7153	0.7181
XGAUC	MLP	0.6521	0.5925	0.5335	0.6617	0.6641	0.6549	0.6645	0.6658	0.6672	0.6717	0.6725
	DCN	0.6630	0.5921	0.4633	0.6600	0.6609	0.6542	0.6641	0.6676	0.6686	0.6716	0.6735
	DCNV2	0.6624	0.5922	0.4687	0.6616	0.6632	0.6542	0.6642	0.6672	0.6671	0.6718	0.6729
	GDCN	0.6578	0.5921	0.4671	0.6612	0.6636	0.6543	0.6646	0.6669	0.6690	0.6722	0.6739

Table 1: Raw watch-time error and accuracy metrics for relative advantage debiasing methods versus baselines.

order of engagement between item pairs; and (3) group-XAUC, which measures whether predicted scores preserve the true watch-time ordering of item pairs within each group. For online A/B tests, we track common engagement metrics such as watch time, finish rate, skip rate, etc. Extensions such as generalized group AUC for user or video cohorts, as well as user clustering procedures, are introduced and discussed in the relevant result sections. Duration binning is also employed, as described earlier in the methods.

Watch-Time Prediction

Our primary hypothesis is that RAD’s bounded, homoskedastic quantile labels can reduce prediction error by stabilizing training and removing systematic biases from watch-time data. To test this, we train models with user-side (RAD-U) and video-side (RAD-V) labels across multiple backbone architectures. After training, predicted quantiles are mapped back to the watch-time domain using empirical inverse CDFs estimated from the training set. Following most standard practice in previous works (Zhan et al. 2022; Lin et al. 2025a), watch time is not truncated at video duration. Because users frequently watch videos past the initial “finish” point of the video, untruncated watch time provides additional preference signals.

Table 1 reports mean absolute error (MAE), XAUC, and XGAUC across all architectures. Both RAD-U and RAD-V outperform all baselines, confirming that quantile-based labels lead to more accurate and robust watch-time prediction. Specifically, RAD achieves lower MAE, stronger ranking performance (XAUC), and superior user-centric ranking (XGAUC) compared to all other methods. These results support our hypothesis that quantile-based, distributionally normalized labels not only improve overall accuracy, but also enhance the model’s ability to capture relative user preferences in a more stable and consistent way.

To further leverage complementary information from user and video perspectives, we average their watch-time predictions after transforming the predicted quantiles back to the watch-time domain using the empirical inverse CDF from the training data. This combined approach (RAD-UV) consistently achieves the best or near-best results for all metrics, outperforming either RAD-U or RAD-V individually. The improvement suggests that each perspective captures differ-

ent aspects of user engagement, and their prediction errors offset each other. By averaging, we reduce individual model biases and variance, leading to more robust, accurate, and user-centric predictions, highlighting the value of integrating both perspectives in debiased modeling.

Relative Preference Modeling

While previous sections focused on reconstructing raw watch time from debiased RAD predictions, our broader goal is to assess how well the proposed methods capture underlying user preferences—that is, whether debiased labels yield more accurate orderings within each user and within each video. To this end, we model RAD labels directly and evaluate whether they preserve the ordering of watch time across both axes. This formulation enables a focused assessment of each model’s ability to capture latent preference structures using RAD labels, independent of engagement magnitude.

Group-level Metrics To assess each model’s ability to capture user–video preferences, we generalize the standard XAUC metric into two group-level variants tailored for evaluating our quantile-based approach.

The original XGAUC evaluates, for each user, how well the predicted watch-time order matches the true watch-time order, after mapping RAD predictions back to the watch-time scale. This score is then averaged across users. In our framework, we directly compare the orderings given by RAD predictions in the quantile space to the actual watch-time order, focusing on scale-free preference modeling.

- **User Group XAUC:** For each user, measures how well the RAD-V prediction’s order matches the true watch-time order across videos.
- **Video Group XAUC:** For each video, measures how well RAD-U predictions can rank users in a consistent way compared to their actual watch times.

These group-based metrics provide a more robust and interpretable evaluation of preference modeling by focusing directly on orderings in the quantile space.

Relative Preference Prediction Table 2 reports XAUC scores computed directly on model-predicted CDF values. Across models, debiasing-based approaches—including

Metric	Backbone	VR	PCR	WLR	WTG	D2Co	D2Q	CWM	CQE	RAD-V CDF	RAD-V CDF	RAD-UV CDF
UGroup XAUC	MLP	0.6909	0.6613	0.6068	0.7047	0.7054	0.6984	0.7050	0.7050	—	0.7105	0.7127
	DCN	0.7017	0.6603	0.5463	0.7044	0.7043	0.6971	0.7065	0.7067	—	0.7092	0.7132
	DCNV2	0.6994	0.6607	0.5551	0.7048	0.7044	0.6975	0.7068	0.706	—	0.7105	0.7124
	GDCN	0.6954	0.6603	0.5521	0.7053	0.7052	0.6971	0.7055	0.7057	—	0.7101	0.7128
VGroup XAUC	MLP	0.6389	0.6748	0.5172	0.6752	0.6739	0.6755	0.6713	0.6728	0.6803	—	0.6786
	DCN	0.6599	0.6751	0.4592	0.6725	0.6722	0.6747	0.6716	0.6742	0.6793	—	0.6783
	DCNV2	0.6533	0.6749	0.4749	0.6754	0.6749	0.6745	0.6716	0.6736	0.6785	—	0.6776
	GDCN	0.6306	0.6753	0.4853	0.6749	0.6747	0.6743	0.6700	0.6728	0.6805	—	0.6771

Table 2: User and video grouped XAUC metrics. In relative advantage debiasing methods, XAUC is calculated directly based on CDF values. The combined RAD method, RAD-UV, estimates a joint CDF computed via Bayesian evidence fusion.

D2Q, CWM, and all RAD variants—consistently outperform classical baselines such as VR, PCR, and WLR. This highlights the importance of explicitly correcting for confounding effects in modeling user engagement and preference. Among all debiasing methods, RAD achieves the strongest overall performance: RAD-U achieves the highest scores on User Group XAUC, while RAD-V leads on Video Group XAUC, supporting our hypothesis that side-specific debiasing improves intra-group preference modeling.

Notably, the fused RAD-UV model, which combines user- and video-side predictions in the latent space, matches or exceeds the performance of the single-sided models and all baselines on group XAUC (Table 2). This demonstrates a clear benefit of jointly modeling both sides for a balanced and robust performance. Overall, these findings demonstrate that quantile-based debiasing with RAD, particularly when integrating different umbrella factors, yields the most accurate and robust preference rankings.

Method	Wasserstein Distance
CQE + MLP	9.3945
CQE + DCN	9.0804
CQE + DCNV2	9.2101
CQE + GDCN	9.1529
MQ + MLP	2.2793
Min Wasserstein Distance	2.1500

Table 3: Average Wasserstein distance per estimated user group distribution, comparing CQE with MLP based multiquantile estimation.

Two-Stage Architecture Ablations

RAD uses umbrella factors to debias a wide range of confounding signals, outperforming single-factor methods such as D2Q. Interestingly, RAD also surpasses CQE, although both leverage watch-time distributions for preference learning. The key distinction is that RAD explicitly separates distribution estimation from preference modeling, whereas CQE entangles both in a single joint process. In this ablation, we address two key objectives: (1) evaluating whether RAD’s two-stage design offers advantages over the one-stage CQE architecture, and (2) testing whether costly per-ID empirical quantiles can be replaced by a compact, learnable distributional embedding without sacrificing accuracy

XGAUC	CQE	MQ + RAD-U	RAD-U
MLP	0.7044	0.7101	0.7131
DCN	0.7071	0.7105	0.7136
DCNV2	0.7066	0.7099	0.7128
GDCN	0.7059	0.7100	0.7130

Table 4: RAD with multiquantile estimation of the user-side distribution (RAD-U + MQ) consistently outperforms CQE, and closely matches performance of RAD-U with exact CDF labels.

or overall model performance.

Learnable Distribution Embedding In online deployment, converting watch times to RAD labels typically requires storing and querying large watch-time histories, which adds overhead from additional indexing services and increases latency. To address this, we propose a distributional embedding approach that learns the distributional information directly as neural network parameters, thereby eliminating the need for historical data retrieval. Each umbrella factor (e.g., user or video cohort) is mapped to a learned embedding, which is then processed by a shared MLP to produce $K = 100$ raw logits. After a ReLU activation and cumulative sum, these logits form a non-negative, strictly increasing sequence of quantile estimates. The model is trained using quantile regression to align the predicted quantiles with empirical watch-time quantiles. At inference, this setup enables the quantile function to be generated without external lookup. In our ablation studies, we systematically compare the accuracy of this learnable embedding approach with baseline methods, including empirical quantile estimation and CQE.

Mitigating Sparsity with User Clustering Accurately learning user- or video-side distributional information through quantile embeddings requires robust and statistically supported quantile estimates. However, in the KuaiRand dataset, neither individual user nor video cohorts offer enough samples for stable quantile learning. For example, the lowest 10th percentile of user and video cohorts—ranked by sample size—contains no more than ten samples per cohort (see Appendix), raising concerns about the reliability of quantile embedding learning and its evaluation. To mitigate this potential issue in our ablation anal-

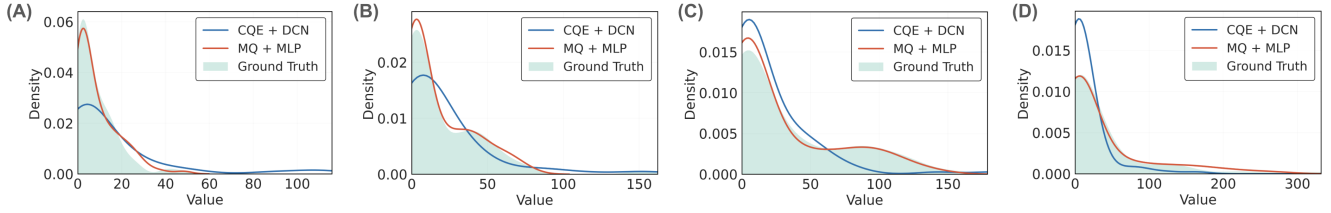


Figure 2: Gaussian kernel density estimates of predicted versus ground-truth watch-time distributions for user-side clusters grouped by training-set size quartile: (A) bottom 25 %, (B) 25–50 %, (C) 50–75 %, and (D) top 25 %. CQE’s single-stage estimates (blue) fail to capture the lower-value regions in (A) and (D), miss secondary modes in (B) and (C), and underestimate heavy tails in (A), (B), and (D), leading to larger errors. In contrast, the MQ + MLP multiquantile model (red) more closely follows the ground-truth curves (green) across all cohort sizes, effectively modeling both typical and irregular shapes.

Methods	CQE	RAD-U	RAD-V	RAD-UV
User Group XAUC	0.623	0.642	—	0.637
Video Group XAUC	0.594	—	0.631	0.625

Table 5: Group XAUC for offline industrial dataset.

Metric	RAD-U	RAD-V	RAD-UV
Active Days	0.0225%	0.0160%	0.0290%
Watch Time	0.2530%	0.0562%	0.3246%
Watch Count	0.0451%	0.2554%	0.0744%
Finish Playing	1.3296%	0.7531%	1.1125%
Skip Rate ↓	-0.4347%	-0.1887%	-0.8640%

Table 6: Metrics for online experiments with relative advantage debiasing

yses, we cluster users into ten groups using K-modes clustering (Chaturvedi, Green, and Carroll 2001) on sparse user features (see Appendix), treating each group as a clustered analogue of the user ID. This approach helps ensure each cohort has sufficient samples for robust quantile estimation and embedding training. In our ablation studies, these cluster IDs serve as the umbrella condition for deriving user-side RAD labels.

Two-stage Benefits and Embedding Efficiency To evaluate the two-stage architecture and distributional embeddings, we compare (1) RAD-U with empirical quantiles (two-stage), (2) RAD-U with learnable distribution embeddings (two-stage + embeddings), and (3) CQE (one-stage).

We first assess distribution matching quality using the 1-Wasserstein distance (Table 3), where the empirical-quantile baseline (no learning) sets the minimum achievable error between training and test sets. For CQE, predicted user–video distributions are aggregated into a user-side cohort distribution via Monte Carlo mixture sampling (Neal 1992). RAD-U’s learned embeddings closely approach this lower bound and outperform CQE, demonstrating that separating CDF estimation from preference learning produces more accurate distribution fits without requiring storage of watch-time logs. Figure 2 provides representative examples, showing that RAD-U’s multiquantile embeddings consistently align with the ground truth for various user group sizes, while CQE’s estimates diverge.

To evaluate preference modeling, we compare methods using User Group XAUC (Table 4). Across all architectures, both RAD-based approaches, either with empirical quantiles or learned distribution embeddings, consistently outperform the one-stage CQE method. This further underscores the advantage of the two-stage design observed in distribution matching. We also report MAE and XAUC for watch-time prediction, where both two-stage methods consistently surpass CQE, with the embedding-based approach achieving results close to those of empirical quantiles. Taken together, these findings highlight the robustness and accuracy of the two-stage architecture. At the same time, distributional embeddings deliver near equivalent predictive performance while providing greater efficiency and scalability, making them more practical for real-world deployment.

Offline Industrial Dataset We also evaluate RAD and CQE on a large-scale offline industrial dataset. In this setting, RAD employs distributional embeddings to estimate quantiles for real-world systems. Models are trained on 14 days of data and evaluated on a held-out day. With ample samples per umbrella factor, group XAUC is computed directly for both user-side (RAD-U) and video-side (RAD-V) without clustering. As shown in Table 5, RAD consistently and substantially outperforms CQE on both user-group and video-group XAUC, validating its robustness in high-data scenarios for industrial recommendation.

Online A/B Experiments

We further evaluate the RAD methods in a large-scale short-video recommendation platform. In this setting, the abundance of data allows us to test robustness and effectiveness in high-data regimes and real-world settings.

Implementation Details We directly learn and deploy three RAD-based models, RAD-U, RAD-V, and their Bayesian fusion (RAD-UV), as candidate rankers for user–item preference. The current production model serves as the baseline. For online evaluation, we compare the outcomes of these three new models against the baseline, forming four experimental groups. Each group is assigned an equal portion of platform traffic, ensuring fair comparison under consistent conditions.

Results Table 6 reports the results of large-scale online experiments. RAD-U produces the largest gain in Finish Playing, while RAD-V achieves the highest result for Watch Count, reflecting the complementary advantages of user- and video-side debiasing. All RAD-based methods improve user engagement over the baseline, with the combined RAD-UV model showing the most balanced gains, leading in Active Days, Watch Time, and Skip Rate. These findings highlight both the overall effectiveness of RAD-based debiasing and the added value of integrating user- and video-side signals for real-world recommendation.

Conclusion

We have presented Relative Advantage Debiasing (RAD), a principled framework for mitigating confounding effects in watch-time-based recommendation. By reframing raw watch time as a cohort-relative quantile, RAD systematically and jointly corrects both user- and item-side biases through the utilization and integration of umbrella factors. Our learnable distributional embedding enables efficient, scalable deployment without the need for historical data storage. Both offline and online experiments demonstrate that RAD consistently improves watch-time prediction and ranking quality over existing baselines, with the hybrid RAD showing most balanced gains across key engagement metrics.

RAD opens several promising directions for the broader recommendation and learning community. For example, RAD quantiles make it possible to jointly optimize watch time with various interaction signals (such as likes and comments) within a normalized space, thus providing a natural foundation for multi-task learning (Lu, Dong, and Smyth 2018). In the context of reinforcement learning, quantile-based rewards can help stabilize policy optimization and readily connect to reward transformation frameworks such as Group Relative Policy Optimization (GRPO) (Shao et al. 2024). These normalized signals may also enhance generative listwise recommendation (Liu et al. 2023), supplying calibrated feedback for more effective credit assignment in sequence or slate-based models (Bello et al. 2018). Moreover, since quantile labels inherently adjust for duration and popularity, future work should investigate their effects on fairness, robustness, and cold-start performance (Zhu et al. 2021), as well as their applicability to other domains—including music, news, and e-commerce (Chen and Lee 2024).

Acknowledgments

We appreciate Kunpeng Liu and Danhui Guan for foundational work and helpful discussions that refined the method.

References

- Angrist, J.; Chernozhukov, V.; and Fernandez-Val, I. 2004. Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure. Working Paper 10428, National Bureau of Economic Research.
- Bae, H.-K.; Lee, Y.-C.; Han, K.; and Kim, S.-W. 2023. A Competition-Aware Approach to Accurate TV Show Recommendation. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 2822–2834.
- Bello, I.; Kulkarni, S.; Jain, S.; Boutilier, C.; Chi, E.; Eban, E.; Luo, X.; Mackey, A.; and Meshi, O. 2018. Seq2Slate: Re-ranking and slate optimization with RNNs. *arXiv preprint arXiv:1810.02019*.
- Cade, B.; and Noon, B. 2003. A Gentle Introduction to Quantile Regression for Ecologists. *Frontiers in Ecology and the Environment*, 1: 412–420.
- Chaturvedi, A.; Green, P. E.; and Carroll, J. D. 2001. K-modes clustering. *Journal of classification*, 18: 35–55.
- Chen, Y.-C.; and Lee, W.-C. 2024. A novel cross-domain recommendation with evolution learning. *ACM Transactions on Internet Technology*, 24(1): 1–23.
- Covington, P.; Adams, J.; and Sargin, E. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16*, 191–198. New York, NY, USA: Association for Computing Machinery. ISBN 9781450340359.
- Dabney, W.; Rowland, M.; Bellemare, M. G.; and Munos, R. 2018. Distributional reinforcement learning with quantile regression. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'18/IAAI'18/EAAI'18*. AAAI Press. ISBN 978-1-57735-800-8.
- Davidson, J.; Liebald, B.; Liu, J.; Nandy, P.; Van Vleet, T.; Gargi, U.; Gupta, S.; He, Y.; Lambert, M.; Livingston, B.; and Sampath, D. 2010. The YouTube video recommendation system. In *Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10*, 293–296. New York, NY, USA: Association for Computing Machinery. ISBN 9781605589060.
- Dorka, N. 2024. Quantile Regression for Distributional Reward Models in RLHF.
- Gao, C.; Li, S.; Zhang, Y.; Chen, J.; Li, B.; Lei, W.; Jiang, P.; and He, X. 2022. KuaiRand: An Unbiased Sequential Recommendation Dataset with Randomly Exposed Videos. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, CIKM '22*, 3953–3957. New York, NY, USA: Association for Computing Machinery. ISBN 9781450392365.
- Järvelin, K.; and Kekäläinen, J. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.*, 20(4): 422–446.
- Jing, P.; Liu, X.; Zhang, L.; Li, Y.; Liu, Y.; and Su, Y. 2024. Multimodal Attentive Representation Learning for Micro-video Multi-label Classification. *ACM Trans. Multimedia Comput. Commun. Appl.*, 20(6).
- Koenker, R.; and Bassett, G. 1978. Regression Quantiles. *Econometrica*, 46(1): 33–50.
- Lin, C.; Liu, S.; Wang, C.; and Liu, Y. 2025a. Conditional Quantile Estimation for Uncertain Watch Time in Short-Video Recommendation. arXiv:2407.12223.

- Lin, C.; Wang, C.; Xie, A.; Wang, W.; Zhang, Z.; Ruan, C.; Huang, Y.; and Liu, Y. 2025b. AlignPxr: Aligning Predicted Behavior Distributions for Bias-Free Video Recommendations. *arXiv:2503.06920*.
- Lin, X.; Chen, X.; Song, L.; Liu, J.; Li, B.; and Jiang, P. 2023. Tree based Progressive Regression Model for Watch-Time Prediction in Short-video Recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '23*, 4497–4506. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701030.
- Liu, S.; Cai, Q.; He, Z.; Sun, B.; McAuley, J.; Zheng, D.; Jiang, P.; and Gai, K. 2023. Generative flow network for listwise recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1524–1534.
- Liu, S.; Chen, Z.; Liu, H.; and Hu, X. 2019. User-Video Co-Attention Network for Personalized Micro-video Recommendation. In *The World Wide Web Conference, WWW '19*, 3020–3026. New York, NY, USA: Association for Computing Machinery. ISBN 9781450366748.
- Liu, Y.; Liu, Q.; Tian, Y.; Wang, C.; Niu, Y.; Song, Y.; and Li, C. 2021. Concept-Aware Denoising Graph Neural Network for Micro-Video Recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management, CIKM '21*, 1099–1108. New York, NY, USA: Association for Computing Machinery. ISBN 9781450384469.
- Lu, Y.; Dong, R.; and Smyth, B. 2018. Why I like it: multi-task learning for recommendation and explanation. In *Proceedings of the 12th ACM Conference on Recommender Systems*, 4–12.
- Meinshausen, N. 2006. Quantile Regression Forests. *Journal of Machine Learning Research*, 7(35): 983–999.
- Neal, R. M. 1992. Bayesian mixture modeling. In *Maximum Entropy and Bayesian Methods: Seattle, 1991*, 197–211. Springer.
- Niu, Z.; Zhou, M.; Wang, L.; Gao, X.; and Hua, G. 2016. Ordinal Regression with Multiple Output CNN for Age Estimation. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4920–4928.
- Padilla, O. H. M.; Tansey, W.; and Chen, Y. 2022. Quantile regression with ReLU Networks: Estimators and minimax rates. *Journal of Machine Learning Research*, 23(247): 1–42.
- Petneházi, G. 2021. Quantile convolutional neural networks for Value at Risk forecasting. *Machine Learning with Applications*, 6: 100096.
- Qin, J.; Zhu, J.; Liu, Y.; Gao, J.; Ying, J.; Liu, C.; Wang, D.; Feng, J.; Deng, C.; Wang, X.; Jiang, J.; Liu, C.; Yu, Y.; Zeng, H.; and Zhang, W. 2023. Learning to Distinguish Multi-User Coupling Behaviors for TV Recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, WSDM '23*, 204–212. New York, NY, USA: Association for Computing Machinery. ISBN 9781450394079.
- Quan, Y.; Ding, J.; Gao, C.; Li, N.; Yi, L.; Jin, D.; and Li, Y. 2023. Alleviating Video-length Effect for Micro-video Recommendation. *ACM Trans. Inf. Syst.*, 42(2).
- Rodrigues, F.; and Pereira, F. C. 2020. Beyond Expectation: Deep Joint Mean and Quantile Regression for Spatiotemporal Problems. *IEEE Transactions on Neural Networks and Learning Systems*, 31(12): 5377–5389.
- Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y.; Wu, Y.; et al. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.
- Sun, J.; Ding, Z.; Chen, X.; Chen, Q.; Wang, Y.; Zhan, K.; and Wang, B. 2024. CREAD: A Classification-Restoration Framework with Error Adaptive Discretization for Watch Time Prediction in Video Recommender Systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8): 9027–9034.
- Tang, S.; Li, Q.; Wang, D.; Gao, C.; Xiao, W.; Zhao, D.; Jiang, Y.; Ma, Q.; and Zhang, A. 2023. Counterfactual video recommendation for duration debiasing. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4894–4903.
- Tang, W.; Shen, G.; Lin, Y.; and Huang, J. 2022. Nonparametric Quantile Regression: Non-Crossing Constraints and Conformal Prediction.
- Wang, F.; Gu, H.; Li, D.; Lu, T.; Zhang, P.; and Gu, N. 2023. Towards Deeper, Lighter and Interpretable Cross Network for CTR Prediction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM '23*, 2523–2533. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701245.
- Wang, R.; Fu, B.; Fu, G.; and Wang, M. 2017. Deep & Cross Network for Ad Click Predictions. In *Proceedings of the ADKDD'17, ADKDD'17*. New York, NY, USA: Association for Computing Machinery. ISBN 9781450351942.
- Wang, R.; Shivanna, R.; Cheng, D.; Jain, S.; Lin, D.; Hong, L.; and Chi, E. 2021a. DCN V2: Improved Deep & Cross Network and Practical Lessons for Web-scale Learning to Rank Systems. In *Proceedings of the Web Conference 2021, WWW '21*, 1785–1797. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383127.
- Wang, W.; Feng, F.; He, X.; Nie, L.; and Chua, T.-S. 2021b. Denoising Implicit Feedback for Recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM '21*, 373–381. New York, NY, USA: Association for Computing Machinery. ISBN 9781450382977.
- Yang, S.; Yang, H.; Du, L.; Ganesh, A.; Peng, B.; Liu, B.; Li, S.; and Liu, J. 2025. SWaT: Statistical Modeling of Video Watch Time through User Behavior Analysis. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1, KDD '25*, 2768–2778. New York, NY, USA: Association for Computing Machinery. ISBN 9798400712456.
- Zhan, R.; Pei, C.; Su, Q.; Wen, J.; Wang, X.; Mu, G.; Zheng, D.; Jiang, P.; and Gai, K. 2022. Deconfounding duration bias

in watch-time prediction for video recommendation. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 4472–4481.

Zhang, Y.; Bai, Y.; Chang, J.; Zang, X.; Lu, S.; Lu, J.; Feng, F.; Niu, Y.; and Song, Y. 2023. Leveraging Watch-time Feedback for Short-Video Recommendations: A Causal Labeling Framework. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM '23*, 4952–4959. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701245.

Zhao, H.; Cai, G.; Zhu, J.; Dong, Z.; Xu, J.; and Wen, J.-R. 2024. Counteracting Duration Bias in Video Recommendation via Counterfactual Watch Time. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4455–4466.

Zhao, H.; Zhang, L.; Xu, J.; Cai, G.; Dong, Z.; and Wen, J.-R. 2023. Uncovering User Interest from Biased and Noised Watch Time in Video Recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems, RecSys '23*, 528–539. New York, NY, USA: Association for Computing Machinery. ISBN 9798400702419.

Zheng, Y.; Gao, C.; Ding, J.; Yi, L.; Jin, D.; Li, Y.; and Wang, M. 2022. Dvr: micro-video recommendation optimizing watch-time-gain under duration bias. In *Proceedings of the 30th ACM International Conference on Multimedia*, 334–345.

Zhu, H.; Jin, J.; Tan, C.; Pan, F.; Zeng, Y.; Li, H.; and Gai, K. 2017. Optimized Cost per Click in Taobao Display Advertising. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17*, 2191–2200. New York, NY, USA: Association for Computing Machinery. ISBN 9781450348874.

Zhu, Z.; Kim, J.; Nguyen, T.; Fenton, A.; and Caverlee, J. 2021. Fairness among new items in cold start recommender systems. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 767–776.