

Uncertainty as Remedy: Mitigating Satisfaction Label Bias in Short Video Multi-Objective Ensemble Ranking

Zonghe Shao*
Kuaishou Technology
Beijing, China
zhshao@hust.edu.cn

Tiantian He*
Kuaishou Technology
Beijing, China
hetiantian05@kuaishou.com

Xiaoxiao Xu*[†]
Kuaishou Technology
Beijing, China
xuxiaoxiao05@kuaishou.com

Jiaqi Yu
Kuaishou Technology
Beijing, China
yujiaqi03@kuaishou.com

Minzhi Xie
Kuaishou Technology
Beijing, China
xieminzhi@kuaishou.com

Jinfang Gu
Kuaishou Technology
Beijing, China
gujinfang@kuaishou.com

Yongqi Liu
Kuaishou Technology
Beijing, China
liuyongqi@kuaishou.com

Kaiqiao Zhan
Kuaishou Technology
Beijing, China
zhankaiqiao@kuaishou.com

Kun Gai
Unaffiliated
Beijing, China
gai.kun@qq.com

Abstract

The core objective of short video recommendation is to model users' unobservable true satisfaction with recommended videos. As the dominant industrial framework, end-to-end multi-objective ensemble ranking models are typically trained with multi-dimensional dense user behavioral signals, such as clicks and watch time. However, these behavioral signals are partial, fragmented, and often mutually conflicting user satisfaction proxies, introducing uncertainty and label bias into satisfaction modeling. Conventional deterministic models overlook this uncertainty, which may exacerbate satisfaction label bias and result in suboptimal convergence. This paper proposes **UAME**, an **Uncertainty-Aware end-to-end Multi-objective Ensemble ranking framework** for short video recommendation. UAME represents the model's prediction as a Gaussian scoring variable, where the mean denotes the predicted satisfaction score and the variance quantifies predictive uncertainty associated with this score. We further design a probabilistic pairwise ranking loss, and construct an uncertainty-aware sample-level weighting scheme to mitigate the bias. We further provide theoretical analysis suggesting that the weighting scheme helps mitigate satisfaction label bias. Extensive offline and online experiments on a large-scale industrial short video platform demonstrate that UAME consistently improves two state-of-the-art paradigms, EMER and EASQ, and better aligns with questionnaire-based user satisfaction. UAME

has been deployed in our production short-video recommendation system and continues to deliver stable, statistically significant gains.

CCS Concepts

• **Information systems** → **Recommender systems**.

Keywords

Video Recommendation, Ensemble Ranking, User Satisfaction, Uncertainty

ACM Reference Format:

Zonghe Shao, Tiantian He, Xiaoxiao Xu, Jiaqi Yu, Minzhi Xie, Jinfang Gu, Yongqi Liu, Kaiqiao Zhan, and Kun Gai. 2026. Uncertainty as Remedy: Mitigating Satisfaction Label Bias in Short Video Multi-Objective Ensemble Ranking. In . ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 Introduction

Short-video recommendation platforms serve hundreds of millions of daily active users (DAUs), where the core goal of the recommender system is to accurately model the user's unobservable true satisfaction with content [29]. To model this latent variable, the dominant industrial paradigm is multi-objective ranking, which typically follows a two-stage architecture [23], as shown in Figures 1a and 1b. Firstly, large-scale models are employed to estimate user satisfaction in multiple dimensions, e.g., click, watch time, like, collectively referred to as pxtr scores [14]. Secondly, these pxtr scores are fused into a unified ranking score. Early approaches rely on manually designed heuristic formulas, e.g., weighted sum or product. Subsequent works have improved fusion by learning the combination weights via various methods, including cross-entropy methods [2, 22], grid search [35], and reinforcement learning [30, 40], yet they still lack end-to-end personalized ranking.

To address these bottlenecks, end-to-end multi-objective ensemble ranking [1, 7, 16, 36, 37] has emerged as a promising direction. This paradigm replaces manually designed heuristic fusion formula with deep learning models, and uses multiple pxtr signals as user

*These authors contributed equally to this work.

[†]The corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference'17, Washington, DC, USA

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

satisfaction proxies, enabling end-to-end ranking. Representative works such as EMER [7], Pantheon [1], and HarmonRank [36] have validated the effectiveness of this paradigm in large-scale industrial short-video recommendation scenarios, and achieved substantial performance gains over heuristic formulas. Despite the progress, they still exhibit two important limitations that prevent alignment with users' real comprehensive satisfaction.

On the one hand, they fail to resolve the inherent bias between multi-dimensional satisfaction proxies and true user satisfaction [16]. Existing methods implicitly treat pxtr signals as sufficiently reliable user satisfaction proxies and adopt an equal-weight loss for optimization. In practice, heterogeneous pxtr signals only capture fragmented dimensions of short-video user behaviors [31], which carry ranking constraints that often contradict each other and deviate from real comprehensive preference, introducing inherent satisfaction label bias. The equal-weight loss can further aggravate this bias, which makes items with consistent pxtr ranking dominate model updates, while high-conflict items critical to satisfaction modeling are severely under-optimized. On the other hand, the deterministic modeling paradigm cannot reconcile conflicting optimization objectives. Existing models output a deterministic score, assuming equal confidence in all predictions and that a fixed score can satisfy all conflicting pxtrs. Both assumptions are often violated in real-world [3], forcing the model to converge to a compromised local optimum [17] without uncertainty quantification [32].

Additionally, existing uncertainty-aware recommendation works mostly focus on point-wise prediction tasks [4, 26], or mainly use uncertainty as an auxiliary signal for post-hoc ranking adjustment [21, 38], with few attempts to embed it into the core end-to-end optimization pipeline to mitigate label bias. They also rarely clarify the source of uncertainty tightly coupled with satisfaction label bias.

In this paper, we propose the **Uncertainty-Aware end-to-end Multi-objective Ensemble ranking (UAME)** framework. Our core insight is that, under multi-objective ranking labels, the predictive uncertainty driven by multi-objective conflicts is empirically associated with sample-level satisfaction label bias, which reflects the bias between multi-dimensional satisfaction proxies and true user satisfaction. Based on this, we construct an adaptive weighting mechanism solely via predicted uncertainty, which can be embedded into the core optimization pipeline to mitigate satisfaction label bias and align the model with true user satisfaction.

The core contributions of this paper are summarized as follows:

- We introduce an uncertainty-aware scoring formulation for multi-objective ensemble ranking, where the mean estimates the satisfaction score and the variance captures multi-objective conflict-driven predictive uncertainty.
- We develop a probabilistic pairwise ranking objective with uncertainty-aware sample weighting, which leverages learned uncertainty to emphasize high-conflict item pairs and mitigate satisfaction label bias during training.
- We deploy and extensively validate the proposed framework on a large-scale industrial short-video recommendation system with hundreds of millions of DAUs. Extensive offline

experiments, online A/B tests, and questionnaire-based analyses demonstrate its effectiveness, stability, and improved alignment with true user satisfaction.

2 Related Works

2.1 End-to-end Multi-objective Ensemble Ranking

End-to-end multi-objective ensemble ranking has become the mainstream research direction of industrial short-video recommendation, which replaces the manual fusion formula in the traditional two-stage paradigm with joint optimization of multi-objective pxtr signals. Representative works such as EMER [7], Pantheon [1], and HarmonRank [36] have validated the effectiveness of this paradigm in large-scale industrial recommendation. Despite their effectiveness in eliminating manual fusion constraints, all these methods share a core limitation that they rely heavily on user behavior-derived pxtr signals as ranking labels, thus introducing the inherent bias between multi-dimensional user satisfaction proxies and users' true comprehensive satisfaction. Existing works have attempted to resolve this bias with additional supervision. For example, EASQ [16] introduced explicit satisfaction feedback from user questionnaires as high-quality supervision to align the ranking model with user satisfaction. Different from this line of works, our work analyzes the connection between the aforementioned bias and model prediction uncertainty, and proposes an uncertainty-aware adaptive weighting scheme for item pairs, which mitigates the label bias in the core end-to-end training pipeline without relying on additional supervision signals beyond the existing pxtr behavior labels.

2.2 Uncertainty Estimation in RecSys

Classical machine learning theory divides uncertainty into epistemic uncertainty and aleatoric uncertainty [10], and this work mainly discusses aleatoric uncertainty. Early works extensively explored methods to predict uncertainty for point-wise tasks such as click-through rate (CTR) prediction [19], where uncertainty is typically attributed to label noise, e.g., accidental clicks [9, 20] and distribution shift [13, 25]. These studies mainly use uncertainty to characterize prediction confidence in point-wise recommendation tasks. However, end-to-end multi-objective ensemble ranking [7], which relies on multiple heterogeneous pxtr labels for supervision, introduces a new source of uncertainty in modeling. While label noise and distribution shift still exist, the dominant source of uncertainty here is intrinsic conflicts between multiple behavior pxtr labels for the same user-item pair. A closely related line of research is uncertainty-aware multi-task learning (MTL) [11], yet these methods focus on task-level loss balancing for point-wise prediction tasks to alleviate global conflicts across different tasks [18, 33, 34]. In contrast, our work targets sample-level satisfaction label bias in ranking caused by inconsistent multi-objective behavior labels, and models uncertainty to mitigate bias, which is fundamentally different from task-centric design of MTL and remains unexplored in multi-objective ensemble ranking.

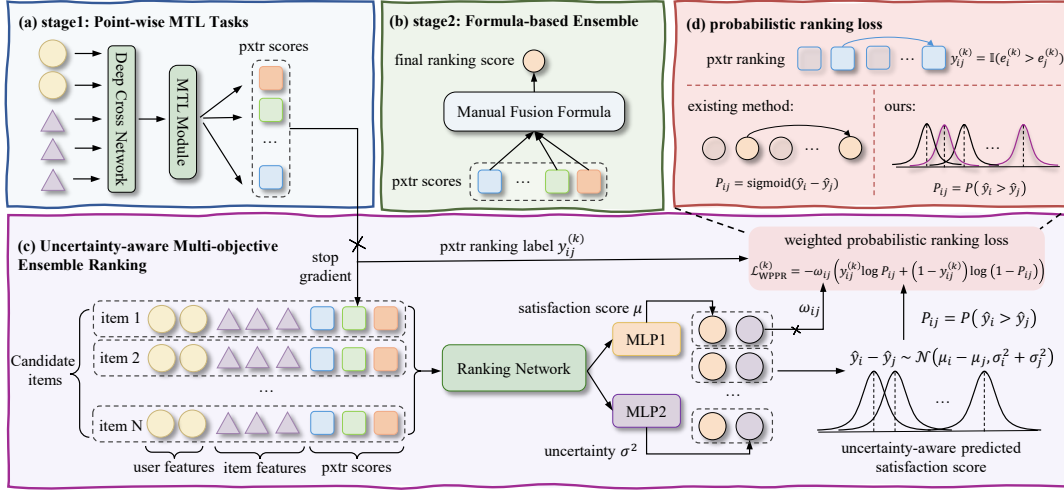


Figure 1: Overview of multi-objective ensemble ranking. (a) Stage 1: Point-wise MTL module generates pxtr scores for different user behaviors. (b) Stage 2: Conventional manual formula-based ensemble of pxtr scores. (c) The proposed uncertainty-aware multi-objective ensemble ranking framework. (d) Comparison of deterministic and our probabilistic pairwise ranking loss.

2.3 Uncertainty in Ranking

Ranking lies at the core of recommendation system, aiming to align users' true satisfaction via optimal item lists. Inherent prediction uncertainty from label noise, behavior bias [39] and distribution shift [8] has been addressed by existing works using uncertainty as a core ranking optimization signal. Other works [38] have alleviated the exploitation bias in learning-to-rank via an uncertainty-aware empirical Bayesian method, and [27] have modeled uncertainty in the top-K ranking problem based on score distribution modeling. However, existing research on uncertainty for ranking models still has a core paradigm limitation. Most existing works treat uncertainty as an auxiliary signal to correct the predicted ranking scores [12, 15] and perform post-hoc ranking adjustment [21]. Such post-hoc usage does not directly mitigate the label bias in the core optimization process. Unlike these existing works, our method embeds uncertainty into the core training pipeline to directly mitigate the label bias during optimization, while satisfying the low-latency requirements of industrial applications.

3 Method

In this section, we present the core framework of **UAME**, including satisfaction-uncertainty modeling, probabilistic pairwise ranking loss, and uncertainty-aware adaptive weighting mechanism, followed by theoretical analysis.

3.1 Problem Statement

We formally define the end-to-end multi-objective ensemble ranking task as follows. Let U denote the user set, V the short-video item set, and c the context features. For a user $u \in U$ under context c , the model takes the candidate set $I_{cand}(u, c)$ as input, and aims to generate a ranked list that maximizes the user's unobservable true satisfaction s_i for each candidate item v_i . The critical challenge of this task is that s_i is a latent variable unavailable for direct model

supervision. Therefore, existing methods rely on behavior-based pxtr signals as multi-dimensional user satisfaction proxies, which are generated by pre-trained MTL models for core interaction objectives such as click, watch time, like, comment, and follow. For item v_i , its K -dimensional pxtr scores are denoted as $\{e_i^{(k)}\}_{k=1}^K$. For any item pair (v_i, v_j) , the ideal pairwise ranking label is $y_{ij}^* = \mathbb{I}\{s_i > s_j\}$, and the pairwise ranking label from the k -th pxtr signal is $y_{ij}^{(k)} = \mathbb{I}\{e_i^{(k)} > e_j^{(k)}\}$, where $\mathbb{I}\{\cdot\}$ is the indicator function. Notably, pxtr signals are inherently biased proxies for user satisfaction, and sample-level conflicts arise across heterogeneous pxtr signals, which are the dominant sources of prediction uncertainty in end-to-end multi-objective ensemble ranking.

3.2 Satisfaction-Uncertainty Modeling

Conventional end-to-end multi-objective ensemble ranking methods output only a single scalar ranking score per candidate item, without quantifying prediction uncertainty. To address this limitation, we design an uncertainty-aware ranking architecture with a weight-sharing two-branch output layer.

As illustrated in Figure 1c, the model takes user features, item features and pxtr scores as input, which are fed into the backbone ranking network. The two output branches fully share the refined feature representation from the backbone ranking network, and generate their respective outputs via two independent lightweight two-layer MLPs. The score prediction branch outputs the model's predicted satisfaction score μ_i , while the uncertainty prediction branch outputs the prediction uncertainty σ_i^2 . Based on the two outputs above, we model the uncertainty-aware scoring variable $\hat{y}_i$ for the i -th candidate item v_i as:

$$\hat{y}_i = \mu_i + \sigma_i \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 1) \quad (1)$$

where μ_i denotes the predicted satisfaction score, and σ_i^2 controls the prediction uncertainty associated with this score.

3.3 Probabilistic Pairwise Loss

3.3.1 Probabilistic Ranking. For a candidate item pair (v_i, v_j) , the core goal of the ranking task is to model the probability that v_i should be displayed prior to v_j . To this end, we define the pairwise ranking probability $P_{i,j}$ as

$$P_{i,j} = P(\hat{y}_i > \hat{y}_j) \quad (2)$$

where $\hat{y}_i$ and $\hat{y}_j$ denote the uncertainty-aware scoring variables of v_i and v_j , respectively.

For tractability, we assume that $\hat{y}_i$ and $\hat{y}_j$ are conditionally independent. Based on equation (1), the pairwise score difference $z = \hat{y}_i - \hat{y}_j$ follows $z \sim \mathcal{N}(\mu_i - \mu_j, \sigma_i^2 + \sigma_j^2)$. From this, we derive the analytical form of the probability that v_i should be displayed prior to v_j , which is the cumulative distribution function (CDF) of the standard Gaussian distribution:

$$P(\hat{y}_i > \hat{y}_j) = P(z > 0) = \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right) \quad (3)$$

where $\Phi(\cdot)$ is the CDF of the standard Gaussian distribution.

3.3.2 Probabilistic Pairwise Ranking Loss. Based on the ranking relation in equation (3), we construct a probabilistic pairwise ranking (PPR) loss to accommodate joint supervision from pxtr signals. For the set of all valid item pairs $\mathcal{D}$, the PPR loss for a pxtr signal is defined as:

$$\mathcal{L}_{\text{PPR}} = - \sum_{(i,j) \in \mathcal{D}} [y_{ij}^{(k)} \cdot \log(P(\hat{y}_i > \hat{y}_j)) + (1 - y_{ij}^{(k)}) \cdot \log(1 - P(\hat{y}_i > \hat{y}_j))] \quad (4)$$

where $y_{ij}^{(k)} \in \{0, 1\}$ is the pairwise ranking label of the k -th pxtr. $y_{ij}^{(k)} = 1$ if v_i ranks prior than v_j , and 0 otherwise.

The loss in equation (4) enables the optimization of the satisfaction score and uncertainty. When the model correctly predicts the ranking of an item pair (i.e., $\mu_i > \mu_j$ with $y_{ij}^{(k)} = 1$), the model tends to increase the mean gap $\mu_i - \mu_j$ and decrease σ_i^2 and σ_j^2 , thereby increasing $\frac{\mu_i - \mu_j}{\sqrt{\sigma_i^2 + \sigma_j^2}}$. This drives $P(\hat{y}_i > \hat{y}_j)$ closer to 1, reduces the loss, and enhances the confidence of the correct ranking decision. Conversely, the model tends to adjust the means toward the correct order, while increasing σ_i^2 and σ_j^2 to reduce the absolute value of $\frac{\mu_i - \mu_j}{\sqrt{\sigma_i^2 + \sigma_j^2}}$, masking the prediction more uncertain.

However, this mechanism introduces a potential risk that the model may take an undesired shortcut by increasing the uncertainty σ_i^2 infinitely for all items to avoid learning the correct ranking of μ_i , which leads to a degenerate behavior. To prevent this, we introduce an uncertainty regularization term [11] that penalizes excessively large uncertainty, guiding the model to maintain a reasonable and discriminative uncertainty distribution:

$$\mathcal{L}_{\text{reg}} = \sum_{(i,j) \in \mathcal{D}} \log(1 + \sigma_i^2 + \sigma_j^2) \quad (5)$$

3.3.3 Auxiliary Constraint Loss. Although the PPR loss can jointly optimize the predicted satisfaction score and its uncertainty, the learned uncertainty is only implicitly supervised and may not explicitly reflect pxtr conflicts. To make the learned uncertainty more explicitly characterize the level of conflict across pxtr signals, we

introduce an auxiliary constraint loss, which enforces alignment between the model's uncertainty output and the level of pxtrs conflict of items. Specifically, for each candidate item, we quantify its level of conflict as the standard deviation of its ranking positions across multiple pxtr signals. Based on this measure, we construct a pairwise training set $\mathcal{B}_{\text{aux}}^+$, which consists exclusively of item pairs (i, j) where item v_i exhibits significantly higher conflict than item v_j . We adopt a pairwise logistic loss to construct the auxiliary constraint loss:

$$\begin{aligned} \mathcal{L}_{\text{aux}} &= - \sum_{(i,j) \in \mathcal{B}_{\text{aux}}^+} \log(P(\sigma_i^2 > \sigma_j^2)) \\ &= - \sum_{(i,j) \in \mathcal{B}_{\text{aux}}^+} \log(\text{sigmoid}(\sigma_i^2 - \sigma_j^2)) \end{aligned} \quad (6)$$

where $P(\sigma_i^2 > \sigma_j^2)$ denotes the probability that the uncertainty of item v_i is higher than that of item v_j .

3.4 Uncertainty-aware Adaptive Weighting

3.4.1 Pairwise Comprehensive Uncertainty. In end-to-end multi-objective ensemble ranking, item pairs are not equally informative. Item pairs with consistent rankings across all pxtr signals are relatively easy to optimize, as their relative order is largely unambiguous. In contrast, item pairs with contradictory rankings across pxtrs are more informative for satisfaction modeling, since their relative order directly affects the final ranking quality. However, conventional pairwise ranking does not capture this heterogeneity, as it assigns equal weights to all item pairs. This may over-emphasize easy pairs while under-emphasizing pairs that are more important for multi-objective ranking.

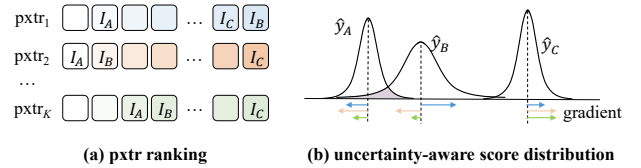


Figure 2: Multi-objective pxtrs ranking conflict and corresponding uncertainty-aware score distribution. (a) Ranking inconsistency of different candidate items under heterogeneous pxtrs. (b) Predicted score distributions with conflict-driven uncertainty.

Figure 2a shows the ranking inconsistency of different candidate items across heterogeneous pxtrs, while Figure 2b visualizes the corresponding optimization gradients and predicted satisfaction distributions. As illustrated in Figure 2, under multi-objective ensemble ranking, the predicted uncertainty naturally differentiates to match this heterogeneity. Items with consistent pxtr rankings such as I_A and I_C exhibit aligned gradients, leading to low-uncertainty and sharp distributions. Items with severe ranking conflicts like I_B suffer from contradictory gradients, leading to high-uncertainty and broad distributions.

Leveraging this property, we quantify the level of pxtrs conflict of each item directly via its predicted uncertainty. For an item pair

(v_i, v_j) , we further define the pairwise comprehensive uncertainty as the sum of the predicted uncertainty of the two items:

$$U_{i,j} = \sigma_i^2 + \sigma_j^2 \quad (7)$$

3.4.2 Adaptive Weighting. To avoid training instability caused by weight distribution, we first apply max-min normalization to the comprehensive uncertainty of all item pairs in the current candidate items, mapping pairwise uncertainty values to $[0, 1]$. We then scale the normalized weights by γ , which aims to ensure that high-uncertainty pairs receive appropriately enhanced weights while maintaining stable convergence of multi-objective joint training. The adaptive weight ω_{ij} for item pair (v_i, v_j) is calculated as:

$$\omega_{ij} = \gamma \times \frac{U_{i,j} - \min(U)}{\max(U) - \min(U)} \quad (8)$$

where γ is the scaling factor, $\min(U)$ and $\max(U)$ denote the minimum and maximum comprehensive uncertainty of all item pairs in the current training batch, respectively.

Building on the adaptive weighting scheme derived above, we integrate the sample-level weight ω_{ij} into the PPR loss, and further combine the auxiliary constraint loss and uncertainty regularization term to form the final loss:

$$\mathcal{L}_{\text{final}} = \sum_{k=1}^K \mathcal{L}_{\text{WPPR}}^{(k)} + \alpha \cdot \mathcal{L}_{\text{reg}} + \beta \cdot \mathcal{L}_{\text{aux}} \quad (9)$$

where $\mathcal{L}_{\text{WPPR}}^{(k)}$ is the weighted probabilistic pairwise ranking loss (WPPR) for the k -th pxtr signal, with the form:

$$\mathcal{L}_{\text{WPPR}}^{(k)} = - \sum_{(i,j) \in \mathcal{D}} \omega_{ij} \cdot \left[y_{ij}^{(k)} \cdot \log(P(\hat{y}_i > \hat{y}_j)) + (1 - y_{ij}^{(k)}) \cdot \log(1 - P(\hat{y}_i > \hat{y}_j)) \right] \quad (10)$$

where K is the number of pxtr signals, $\alpha > 0$ is a hyperparameter that regulates the regularization intensity of the uncertainty term, and $\beta > 0$ is a hyperparameter that balances the optimization trade-off between the weighted ranking loss and the auxiliary constraint loss.

3.5 Theoretical Analysis

In this subsection, we provide the theoretical analysis for uncertainty-aware adaptive weighting mechanism. We first characterize the satisfaction label bias in terms of ranking risk and proxy bias. We then establish the connection among sample-level label bias, pxtrs conflict, and uncertainty. Based on this result, we show that using uncertainty as the adaptive sample weight is a well-motivated way to better align the weighted proxy risk with the true ranking risk.

3.5.1 Ranking Risk Decomposition. The ultimate goal of the multi-objective ensemble ranking model is to model the users' true satisfaction s , i.e., minimize the true population risk over the real-world item pair distribution $\mathcal{P}$, defined as:

$$R(f) = \mathbb{E}_{(i,j) \sim \mathcal{P}} \left[\ell(P_{ij}, y_{ij}^*) \right] \quad (11)$$

where $\ell(\cdot)$ is the pairwise ranking loss.

Existing end-to-end multi-objective ensemble ranking methods only optimize the proxy risk $R_{\text{proxy}}(f)$, which is the average loss

over K pxtr labels:

$$R_{\text{proxy}}(f) = \mathbb{E}_{(i,j) \sim \mathcal{P}} [\hat{\ell}_{ij}], \quad \hat{\ell}_{ij} = \frac{1}{K} \sum_{k=1}^K \ell(P_{ij}, y_{ij}^{(k)}) \quad (12)$$

By the linearity of expectation, the true population risk can be formally decomposed into the following terms:

$$R(f) = R_{\text{proxy}}(f) + \Delta_{\text{proxy}} \quad (13)$$

where $\Delta_{\text{proxy}} = \mathbb{E}_{(i,j) \sim \mathcal{P}} [\ell(P_{ij}, y_{ij}^*) - \hat{\ell}_{ij}]$ is the overall inherent user satisfaction label bias, i.e., proxy bias.

Equation (13) reveals the core limitation of existing methods that they minimize the proxy risk $R_{\text{proxy}}(f)$, but ignore the heterogeneity of sample-level proxy bias $\Delta_{ij} = \ell(P_{ij}, y_{ij}^*) - \hat{\ell}_{ij}$. High-bias items with conflicting pxtr ranking labels can disproportionately affect the optimization outcome, but are assigned equal weights with low-bias items, leading to misalignment with true user satisfaction.

3.5.2 From Proxy Bias to Predictive Uncertainty. To establish the link between sample-level proxy bias and predicted uncertainty, we first analyze how proxy bias arises from the conflict among heterogeneous pxtr signals.

ASSUMPTION 1. The loss $\ell(p, y)$ is L -Lipschitz continuous with respect to soft label y when $p \in [\epsilon, 1 - \epsilon]$, where ϵ is a small constant.

LEMMA 3.1. The sample-level proxy bias satisfies

$$|\Delta_{ij}| = \left| \log \frac{1 - P_{ij}}{P_{ij}} \right| \cdot \left| y_{ij}^* - \frac{1}{K} \sum_{k=1}^K y_{ij}^{(k)} \right| \leq L \cdot \left| y_{ij}^* - \frac{1}{K} \sum_{k=1}^K y_{ij}^{(k)} \right| \quad (14)$$

where $L = \left| \log \frac{1-\epsilon}{\epsilon} \right|$.

Lemma 3.1 shows that the sample-level proxy bias is directly proportional to the mismatch between the true satisfaction label and the averaged pxtr labels. To further connect pxtrs conflict with uncertainty, we introduce the following empirically motivated assumption.

ASSUMPTION 2. Under the optimization of the PPR loss in equation (4), for an item pair (v_i, v_j) , U_{ij} is positively correlated, in expectation, with its level of pxtrs conflict.

This assumption is consistent with the optimization behavior of PPR. Low-conflict pairs receive consistent supervision and can be fitted with a stable mean difference and small variance, whereas high-conflict pairs impose contradictory ranking constraints and therefore tend to induce larger uncertainty. Hence, we represent Proposition 3.2

PROPOSITION 3.2. Under Assumptions 1-2, the sample-level proxy bias $|\Delta_{ij}|$ and the pairwise comprehensive uncertainty U_{ij} are positively associated in expectation through their common dependence on pxtr conflict.

3.5.3 Rationale of Uncertainty-Aware Adaptive Weighting. From the perspective of weighted empirical risk minimization (WERM), we define the weighted proxy risk as

$$R_w(f) = \frac{1}{|\mathcal{D}|} \sum_{(i,j) \in \mathcal{D}} \omega_{ij} \cdot \hat{\ell}_{ij}(f) \quad (15)$$

Proposition 3.2 suggests that the uncertainty U_{ij} is positively associated, in expectation, with the magnitude of proxy bias $|\Delta_{ij}|$. Therefore, using uncertainty as the sample weight encourages WERM to place more emphasis on item pairs that are more likely to incur larger bias [24, 28]. In this sense, uncertainty-aware weighting focuses optimization on high-bias pairs and may improve alignment with the true ranking risk compared to uniform weighting.

4 Experiments

In this section, we conduct extensive offline and online experiments to verify the effectiveness of our proposed method, and answer the following 4 research questions (RQs):

- **RQ1:** Can UAME consistently improve multi-objective ranking across different backbones and pxtr objectives?
- **RQ2:** Can UAME generalize to online user satisfaction-related metrics in the production environment?
- **RQ3:** How do the uncertainty-related components and hyperparameters influence the effectiveness of UAME?
- **RQ4:** Does learned uncertainty reflect satisfaction label bias and improve alignment with user satisfaction?

4.1 Experimental Setup

To systematically verify the effectiveness and superiority of UAME, we conduct comprehensive experiments on large-scale industrial recommendation datasets. The detailed experimental setup is introduced as follows.

4.1.1 Datasets. All experiments are conducted on real-world datasets collected from a large-scale industrial short video recommendation platform with hundreds of millions of DAUs, fully aligning with real-world industrial scenarios for large-scale multi-objective recommendation systems.

4.1.2 Evaluation Metrics. We adopt user Group Area Under the Curve (GAUC) as the primary metric on different key pxtrs to measure the ranking performance of the model, which is more aligned with the core personalized ranking goal of recommendation systems, and can more accurately measure the model's ability to fit each user's personalized preference, compared with the global AUC metric. GAUC is calculated as follows:

$$\text{GAUC} = \frac{\sum_i^{\text{user}} \# \text{impression}_{u_i} \cdot \text{AUC}_{u_i}}{\sum_i \# \text{impression}_{u_i}} \quad (16)$$

where u_i is the i -th user, $\# \text{impression}_{u_i}$ denotes the impression count, and AUC_{u_i} denotes the corresponding AUC value.

4.1.3 Baselines. To evaluate the effectiveness of our method on representative industrial multi-objective ensemble ranking frameworks, we select two state-of-the-art (SOTA) models as our backbone ranking networks:

- **EMER** [7], a widely deployed industrial end-to-end ensemble ranking method.
- **EASQ** [16], a user satisfaction-aligned ranking method, which introduces explicit user questionnaire satisfaction feedback as high-quality ranking supervision.

And we compare our method with four baselines:

- **Base Model** refers to the EMER and EASQ backbone without our uncertainty-aware components.
- **RankDist** [27] models uncertainty by predicting variance, but only uses it to adjust ranking scores during inference.
- **EBRank** [38] adopts empirical Bayes to model prediction uncertainty and utilizes upper confidence bounds (UCB) for exploration.
- **ANSL** [6] assigns optimization weights to samples automatically according to the magnitude of their ranking loss.

Together, these baselines correspond to different settings: using uncertainty only at inference (RankDist, EBRank) or applying sample weighting without uncertainty (ANSL), which highlights the benefit of using uncertainty as sample-level weights during training.

4.1.4 Implementation Details. For fair comparison, all methods are evaluated under the same data split, feature setting, and evaluation protocol. For each baseline, we independently tune its key hyperparameters on the validation set to obtain a strong configuration. For our method, the embedding size of user and item features is fixed at 32, and the embedding dimension of the pxtr is fixed at 8. Training is performed using the Adagrad optimizer [5] on the shuffled samples, while learning rate is set to 0.001. The loss weights α and β in equation (9) are set to 0.02 and 0.1, respectively, and the weight scaling factor γ is set to 2. More implementation and deployment details are provided in Appendix B.

4.2 Overall Performance (RQ1)

To answer **RQ1**, we conduct comparative experiments on the EMER and EASQ backbones. Table 1 presents the offline performance across eight core industrial pxtrs, including **pctr**/ click, **pvtr**/ watch-time, **plvtr**/ long-view, **pcpr**/ complete-view, **pltr**/ like, **pwtr**/ follow, **pcmtr**/ comment, and **pftr**/ share. From these results, we derive the following core conclusions:

(1) Our method achieves overall improvements across both EMER and EASQ backbones, and shows relatively consistent gains on a majority of pxtr metrics. The consistent gains across two structurally distinct backbones also demonstrate the strong generalization ability of our framework, which can be easily embedded into mainstream industrial multi-objective ensemble ranking models without modifying the core backbone structure.

(2) Our method consistently outperforms the original backbones. Compared with the original EMER model, our method brings up to 5.8% relative improvement on core business metrics (e.g., pwtr from 0.684 to 0.724). On the EASQ backbone, the improvement is more prominent, with up to 14.2% relative gain on plvtr (from 0.668 to 0.763) and 9.2% on pctr (from 0.672 to 0.734).

(3) The comparison with baselines fully validates the validity of our core design. RankDist and EBRank, which use uncertainty for inference-stage score adjustment, only achieve partial improvement on a few metrics, suggesting that post-processing may not fully exploit the value of uncertainty. ANSL, which assigns weights purely by loss magnitude, causes performance drops on several key metrics, indicating that generic sample hardness is insufficient for characterizing the label bias in multi-objective ensemble ranking. In contrast, our method prioritizes high-uncertainty pairs via

Table 1: Performance comparisons between UAME and the baselines across two backbones. The best and second-best performance methods are denoted in bold and underlined fonts, respectively.

Methods		pctr	pvtr	plvtr	pcpr	pltr	pwtr	pcmtr	pfttr
EMER [7]	Base Model	<u>0.706</u>	0.667	<u>0.731</u>	0.640	0.688	0.684	0.686	<u>0.706</u>
	RankDist [27]	0.700	0.663	<u>0.729</u>	<u>0.649</u>	0.690	0.677	0.678	0.689
	EBRank [38]	0.686	0.687	0.714	<u>0.623</u>	<u>0.696</u>	0.707	<u>0.697</u>	0.701
	ANSL [6]	0.666	0.715	0.670	0.642	0.669	<u>0.709</u>	0.599	0.692
	UAME	0.715	<u>0.699</u>	0.735	0.657	0.712	0.724	0.711	0.719
EASQ [16]	Base Model	0.672	0.598	0.668	0.631	0.682	<u>0.661</u>	0.703	<u>0.666</u>
	RankDist [27]	<u>0.707</u>	0.578	<u>0.700</u>	0.673	0.673	0.656	0.685	<u>0.650</u>
	EBRank [38]	0.693	0.634	0.689	<u>0.679</u>	0.675	0.650	0.669	0.655
	ANSL [6]	0.670	<u>0.636</u>	0.677	0.632	0.712	0.666	0.683	0.656
	UAME	0.734	0.669	0.763	0.686	<u>0.684</u>	0.675	<u>0.695</u>	0.678

uncertainty-characterized label bias, achieving more stable overall performance without sacrificing major objectives.

4.3 Online A/B Testing (RQ2)

To answer RQ2 and validate the real-world performance of UAME, we conduct online A/B tests over 7 days on an industrial short-video platform. We set up two parallel test groups to verify UAME’s generalization across two representative industrial SOTA backbones including EMER and EASQ, with each group independently allocated 5% of the main feed traffic to enable statistically powered evaluation. The first group uses the fully deployed online EMER model as control, with the experimental group adopting EMER integrated with UAME. The second group uses the user satisfaction-aligned EASQ model as control, with the experimental group adopting EASQ integrated with UAME.

We evaluate a comprehensive set of user satisfaction metrics, covering long-term retention (LT7), implicit consumption behavior (App stay time, watch time, video view count), and explicit positive interaction (Like, follow, forward, comment). As summarized in Table 2, UAME delivers statistically significant improvements across all metrics in both test groups, with the most prominent gains on longviewing and active interaction metrics. Although some absolute gains are numerically small, this is common in mature large-scale recommendation systems, where improvements are typically incremental but consistent across multiple key metrics. These consistent gains indicate that UAME effectively mitigates satisfaction label bias and aligns model optimization with true user satisfaction in large-scale industrial environments.

4.4 Hyperparameter and Ablation Analysis (RQ3)

To answer RQ3 and verify the robustness of our framework, we conduct hyperparameter analysis and ablation analysis on the key components: the regularization coefficient α , the auxiliary loss coefficient β , and the uncertainty normalization scaling factor γ . All experiments are implemented on the EMER backbone with consistent training settings.

4.4.1 Effect of the Regularization Coefficient α . As shown in Figure 3a, we test $\alpha \in \{0, 0.02, 0.05, 0.1\}$, and the performance first

Table 2: Online A/B performance comparison between UAME and the Baseline Method. The p-value is less than 0.005.

Metrics	vs EMER	vs EASQ
LT7	+0.009%	+0.015%
AppStayTime	+0.050%	+0.074%
WatchTime	+0.030%	+0.221%
VideoView	+0.536%	+0.373%
LongView	+1.614%	+1.126%
Like	+0.939%	+1.349%
Follow	+0.641%	+1.299%
Forward	+1.325%	+1.598%
Comment	+0.679%	+0.642%

increases and then decreases as α increases. When α is set to 0.02, our model achieves the best performance. An excessively small α cannot regularize the uncertainty properly, while an excessively large α excessively suppresses the discrimination of uncertainty and weakens the adaptive weighting mechanism.

4.4.2 Effect of the Auxiliary Loss Coefficient β . In Figure 3b, we test $\beta \in \{0, 0.01, 0.1, 0.3\}$, and the model performs optimally when $\beta = 0.1$. A small β leads to poor alignment between uncertainty and the level of real multi-objective pxtrs conflict, whereas a large β makes the auxiliary loss dominate the training and impair the primary ranking task.

4.4.3 Effect of the Uncertainty Scaling Factor γ . Figure 3c illustrates the effect of the scaling factor γ . The model reaches the best performance at $\gamma = 2$. A small γ limits the enhancement of high-uncertainty pairs, while an overly large γ leads to extreme weight distribution and unstable optimization.

4.4.4 Ablation Study. $\alpha = 0$ and $\beta = 0$ in Figures 3a and 3b ablate the uncertainty regularization and auxiliary constraint loss, respectively, while the *uniform* setting in Figure 3c ablates uncertainty-aware weighting by replacing the adaptive weight with a uniform weight, i.e., $\omega_{ij} = 1$. Overall, removing these components leads to inferior performance, confirming that these components all contribute to the effectiveness of UAME.

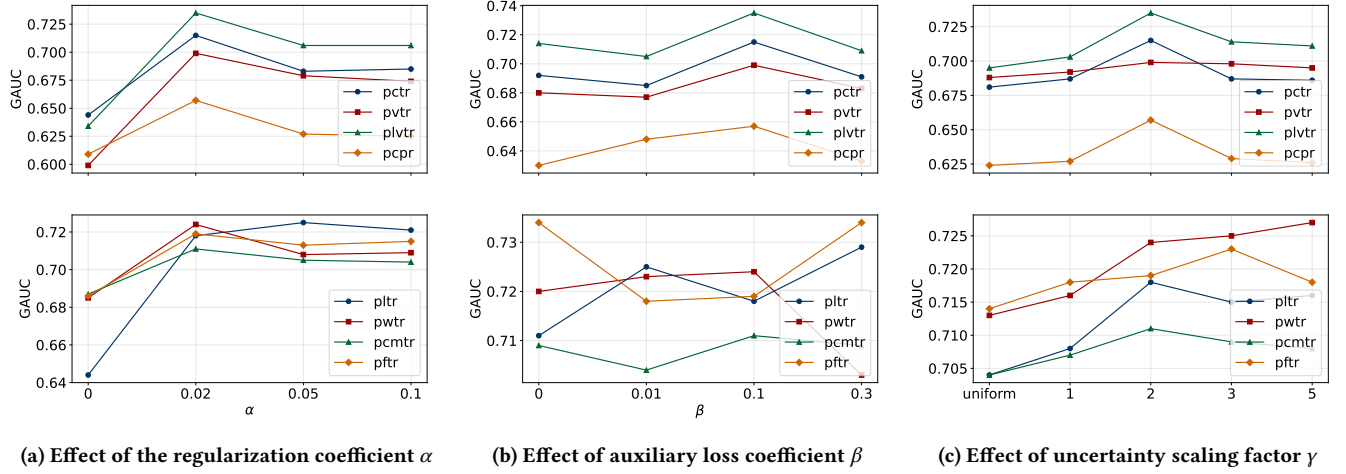


Figure 3: Hyperparameter sensitivity analysis on EMER backbone. (a) Regularization coefficient α , (b) Auxiliary loss coefficient β , (c) Uncertainty-aware weighting and adaptive weight scaling factor γ .

Table 3: User satisfaction modeling performance on EASQ backbone.

Metrics	EASQ	UAME	Improvement $\uparrow$
NDCG@5	0.3753	0.4080	+8.71%
NDCG@10	0.3872	0.4160	+7.44%
HR@1	0.4846	0.5312	+9.62%
HR@5	0.8192	0.9063	+10.63%
HR@10	0.8979	0.9421	+4.92%
MRR	0.6145	0.6675	+8.62%

4.5 Further Analysis (RQ4)

To answer **RQ4**, we further analyze UAME using the questionnaire-based user satisfaction signals on EASQ [16]. Specifically, we study two complementary questions: (i) whether UAME improves alignment with user satisfaction, and (ii) whether the learned uncertainty captures the satisfaction label bias.

4.5.1 True User Satisfaction Modeling Performance. We first evaluate whether UAME improves ranking quality with respect to user satisfaction. Unlike behavior-derived pxtr signals, which carry inherent bias and only capture fragmented dimensions of user preference, user questionnaire feedback more directly reflects users' holistic subjective satisfaction with recommended videos. Using the questionnaire-based signals in EASQ as approximate ground-truth satisfaction and strictly following the experimental settings, we measure ranking performance with standard metrics including Normalized Discounted Cumulative Gain (NDCG), Hit Ratio (HR), and Mean Reciprocal Rank (MRR).

As shown in Table 3, UAME consistently outperforms EASQ across all evaluation metrics, with a maximum relative improvement of 8.71% on NDCG@5 and 10.63% on HR@5, the core metric for top-ranking quality. These results indicate that UAME improves

alignment with user satisfaction beyond optimization on user satisfaction proxies alone.

4.5.2 Relationship between Uncertainty and Satisfaction Label Bias. We next investigate whether the learned uncertainty effectively captures the satisfaction label bias induced by multiple user satisfaction proxies. In industrial recommendation systems, true user satisfaction is typically unobservable, making it difficult to directly assess such bias in standard settings. However, the questionnaire feedback in EASQ provides a reliable approximation of user satisfaction, enabling us to construct a label bias measure that quantifies the inconsistency between the averaged ranking derived from pxtr signals and those based on the questionnaire-based approximate ground-truth. We then study the relationship between the learned uncertainty and the constructed label bias across item pairs. Empirically, on 3 million samples, we observe a positive correlation, with a Pearson correlation coefficient of 0.65 and a Spearman correlation coefficient of 0.67. These findings align with the intuition behind Proposition 3.2 and provides empirical support for Assumption 2.

5 Conclusion

In this work, we address the inherent bias between behavior-derived satisfaction labels and true user satisfaction in end-to-end multi-objective ensemble ranking models for short-video recommendation. We propose UAME, an uncertainty-aware multi-objective ensemble ranking framework. UAME represents the predicted satisfaction score with a Gaussian scoring variable, and uses the uncertainty as an indicator of the level of pxtrs conflict. Based on this, we design an uncertainty-aware adaptive weighting mechanism. Our analysis suggests that this weighting scheme helps mitigate the user satisfaction label bias. Extensive experiments on a large-scale industrial short-video platform with hundreds of millions of DAUs show that UAME achieves consistent, significant improvements over baselines on two representative backbones and enhances alignment with true user satisfaction.

References

- [1] Jiangxia Cao, Pengbo Xu, Yin Cheng, Kaiwei Guo, Jian Tang, Shijun Wang, Dewei Leng, Shuang Yang, Zhaojie Liu, Yanan Niu, et al. 2025. Pantheon: Personalized multi-objective ensemble sort via iterative pareto policy optimization. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. 5575–5582.
- [2] Sirui Chen, Yuan Wang, Zijing Wen, Zhiyu Li, Changshuo Zhang, Xiao Zhang, Quan Lin, Cheng Zhu, and Jun Xu. 2023. Controllable multi-objective re-ranking with policy hypernetworks. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*. 3855–3864.
- [3] Konstantina Christakopoulou, Madeleine Traverse, Trevor Potter, Emma Marriott, Daniel Li, Chris Haulk, Ed H Chi, and Minmin Chen. 2020. Deconfounding user satisfaction estimation from response rate bias. In *Proceedings of the 14th ACM conference on recommender systems*. 450–455.
- [4] Chao Du, Zhifeng Gao, Shuo Yuan, Lining Gao, Ziyang Li, Yifan Zeng, Xiaoqiang Zhu, Jian Xu, Kun Gai, and Kuang-Chih Lee. 2021. Exploration in online advertising systems with deep uncertainty-aware learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2792–2801.
- [5] John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research* 12, 7 (2011).
- [6] Feng Gao, Xin Zhou, Yinning Shao, Yue Wu, Jiahua Gao, Yujian Ren, Fengyang Qi, Ruochen Deng, and Jie Liu. 2025. Both Supply and Precision: Sample Debias and Ranking Consistency Joint Learning for Large Scale Pre-Ranking System. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 11672–11680.
- [7] Tiantian He, Minzhi Xie, Runtong Li, Xiaoxiao Xu, Jiaqi Yu, Zixiu Wang, Lantao Hu, Han Li, and Kun Gai. 2025. An End-to-End Multi-objective Ensemble Ranking Framework for Video Recommendation. *arXiv preprint arXiv:2508.05093* (2025).
- [8] Maria Heuss, Daniel Cohen, Masoud Mansoury, Maarten de Rijke, and Carsten Eickhoff. 2023. Predictive uncertainty-based bias mitigation in ranking. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 762–772.
- [9] Xin Jiang, Kaiqiang Wang, Yinlong Wang, Fengchang Lv, Taiyang Peng, Shuai Yang, Xianteng Wu, Pengye Zhang, Shuo Yuan, and Yifan Zeng. 2024. Deep Uncertainty-Based Explore for Index Construction and Retrieval in Recommendation System. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 4587–4594.
- [10] Alex Kendall and Yarin Gal. 2017. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems* 30 (2017).
- [11] Alex Kendall, Yarin Gal, and Roberto Cipolla. 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7482–7491.
- [12] Norman Knyazev and Harrie Oosterhuis. 2023. A lightweight method for modeling confidence in recommendations with learned beta distributions. In *Proceedings of the 17th ACM conference on recommender systems*. 306–317.
- [13] Ang Li, Jian Hu, Ke Ding, Xiaolu Zhang, Jun Zhou, Yong He, and Xu Min. 2023. Uncertainty-based heterogeneous privileged knowledge distillation for recommendation system. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2471–2475.
- [14] Danwei Li, Zhengyu Zhang, Siyang Yuan, Mingze Gao, Weilin Zhang, Chaofei Yang, Xi Liu, and Jiyang Wang. 2023. Adatt: Adaptive task-to-task fusion network for multitask learning in recommendations. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4370–4379.
- [15] Jiayu Li, Ziyi Ye, Guohao Jian, Zhiqiang Guo, Weizhi Ma, Qingyao Ai, and Min Zhang. 2025. Are Recommenders Self-Aware? Label-Free Recommendation Performance Estimation via Model Uncertainty. *arXiv preprint arXiv:2507.23208* (2025).
- [16] Na Li, Jiaqi Yu, Minzhi Xie, Tiantian He, Xiaoxiao Xu, Zixiu Wang, Lantao Hu, Yongqi Liu, Han Li, Kaiqiao Zhan, et al. 2026. Towards End-to-End Alignment of User Satisfaction via Questionnaire in Video Recommendation. In *Proceedings of the 49th international ACM SIGIR conference on research and development in information retrieval*. To appear.
- [17] Yuyuan Li, Yizhao Zhang, Weiming Liu, Xiaohua Feng, Zhongxuan Han, Chaoshao Chen, and Chenggang Yan. 2025. Multi-objective unlearning in recommender systems via preference guided pareto exploration. *IEEE Transactions on Services Computing* (2025).
- [18] Fuhang Liang, Rucong Xu, and Deng Lin. 2025. Adaprl: Adaptive pairwise regression learning with uncertainty estimation for universal regression tasks. *arXiv preprint arXiv:2501.05809* (2025).
- [19] Huafeng Liu, Jingxuan Wen, Liping Jing, and Jian Yu. 2019. Deep generative ranking for personalized recommendation. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 34–42.
- [20] Taichi Liu, Chen Gao, Zhenyu Wang, Dong Li, Jianye Hao, Depeng Jin, and Yong Li. 2023. Uncertainty-aware consistency learning for cold-start item recommendation. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*. 2466–2470.
- [21] Huanbo Lyu, Daniel Herring, Shiqiao Zhou, Miqing Li, Zheming Zuo, Jelena Ninic, James Andrews, Fabian Spill, and Shuo Wang. 2025. Uncertainty-Aware Dual-Ranking Strategy for Offline Data-Driven Multi-Objective Optimization. *arXiv preprint arXiv:2511.06459* (2025).
- [22] Debabrata Mahapatra, Chaosheng Dong, Yetian Chen, and Michinari Momma. 2023. Multi-label learning to rank through multi-objective optimization. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*. 4605–4616.
- [23] Yue Meng, Cheng Guo, Yi Cao, Tong Liu, and Bo Zheng. 2025. A generative re-ranking model for list-level multi-objective optimization at taobao. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 4213–4218.
- [24] Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. 2013. Learning with noisy labels. *Advances in neural information processing systems* 26 (2013).
- [25] Krishna P Neupane, Ervine Zheng, and Qi Yu. 2024. Evidential stochastic differential equations for time-aware sequential recommendation. *Advances in Neural Information Processing Systems* 37 (2024), 67556–67574.
- [26] Charul Paliwal, Anirban Majumder, and Sivaramakrishnan Kaveri. 2024. Predictive relevance uncertainty for recommendation systems. In *Proceedings of the ACM Web Conference 2024*. 3900–3909.
- [27] Coral Scharf, Carmel Domshlak, Avigdor Gal, and Haggai Roitman. 2025. A Rank-Based Approach to Recommender System’s Top-K Queries with Uncertain Scores. *Proceedings of the ACM on Management of Data* 3, 1 (2025), 1–26.
- [28] Shai Shalev-Shwartz and Shai Ben-David. 2014. *Understanding machine learning: From theory to algorithms*. Cambridge university press.
- [29] Clemencia Siro, Mohammad Aliannejadi, and Maarten De Rijke. 2023. Understanding and predicting user satisfaction with conversational recommender systems. *ACM Transactions on Information Systems* 42, 2 (2023), 1–37.
- [30] Dusan Stamenkovic, Alexandros Karatzoglou, Ioannis Arapakis, Xin Xin, and Kleomenis Katevas. 2022. Choosing the best of both worlds: Diverse and novel recommendations through multi-objective reinforcement learning. In *Proceedings of the fifteenth ACM international conference on web search and data mining*. 957–965.
- [31] Jie Tang, Huiji Gao, Liwei He, and Sanjeev Katariya. 2024. Multi-objective learning to rank by model distillation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5783–5792.
- [32] Hongxu Wang, Zhu Sun, Yingpeng Du, Lu Zhang, Tiantian He, and Yew-Soon Ong. 2025. Uncertain Multi-Objective Recommendation via Orthogonal Meta-Learning Enhanced Bayesian Optimization. *arXiv preprint arXiv:2502.13180* (2025).
- [33] Menghan Wang, Yujie Lin, Guli Lin, Keping Yang, and Xiao-ming Wu. 2020. M2GRL: A multi-task multi-view graph representation learning framework for web-scale recommender systems. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 2349–2358.
- [34] Qichao Wang, Yunhong Lu, Hengyuan Cao, Junyi Zhang, and Min Zhang. 2026. DMGD: Train-Free Dataset Distillation with Semantic-Distribution Matching in Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12417–12427.
- [35] Haolun Wu, Chen Ma, Bhaskar Mitra, Fernando Diaz, and Xue Liu. 2021. Multi-FR: A multi-objective optimization method for achieving two-sided fairness in E-commerce recommendation. *arXiv preprint arXiv:2105.02951* (2021).
- [36] Boyang Xia, Zhou Yu, Zhiliang Zhu, Hanxiao Sun, Biyun Han, Jun Wang, Runnan Liu, and Wenwu Ou. 2026. HarmonRank: Ranking-aligned Multi-objective Ensemble for Live-streaming E-commerce Recommendation. *arXiv preprint arXiv:2601.02955* (2026).
- [37] Zhengrui Xu, Zhe Yang, Zhengxiao Guo, Shukai Liu, Luocheng Lin, Xiaoyan Liu, Yongqi Liu, and Han Li. 2025. UMRE: A Unified Monotonic Transformation for Ranking Ensemble in Recommender Systems. *arXiv preprint arXiv:2508.07613* (2025).
- [38] Tao Yang, Cuize Han, Chen Luo, Parth Gupta, Jeff M Phillips, and Qingyao Ai. 2024. Mitigating exploitation bias in learning to rank with an uncertainty-aware empirical Bayes approach. In *Proceedings of the ACM Web Conference 2024*. 1486–1496.
- [39] Tao Yang, Chen Luo, Hanqing Lu, Parth Gupta, Bing Yin, and Qingyao Ai. 2022. Can clicks be both labels and features? unbiased behavior feature collection and uncertainty-aware learning to rank. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*. 6–17.
- [40] Qihua Zhang, Junning Liu, Yuzhuo Dai, Yiyan Qi, Yifan Yuan, Kunlun Zheng, Fan Huang, and Xianfeng Tan. 2022. Multi-task fusion via reinforcement learning for long-term user satisfaction in recommender systems. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. 4510–4520.

A Proofs of Theoretical Analysis

A.1 Risk Decomposition

Recall that the true ranking risk is defined as

$$R(f) = \mathbb{E}_{(i,j) \sim \mathcal{P}} [\ell(P_{ij}, y_{ij}^*)] \quad (17)$$

and the proxy risk optimized by pxtr ranking labels is

$$R_{\text{proxy}}(f) = \mathbb{E}_{(i,j) \sim \mathcal{P}} [\hat{\ell}_{ij}], \quad \hat{\ell}_{ij} = \frac{1}{K} \sum_{k=1}^K \ell(P_{ij}, y_{ij}^{(k)}) \quad (18)$$

By adding and subtracting $\hat{\ell}_{ij}$ inside the expectation, we have

$$\begin{aligned} R(f) &= \mathbb{E}_{(i,j) \sim \mathcal{P}} [\ell(P_{ij}, y_{ij}^*)] \\ &= \mathbb{E}_{(i,j) \sim \mathcal{P}} [\hat{\ell}_{ij}] + \mathbb{E}_{(i,j) \sim \mathcal{P}} [\ell(P_{ij}, y_{ij}^*) - \hat{\ell}_{ij}] \\ &= R_{\text{proxy}}(f) + \Delta_{\text{proxy}} \end{aligned} \quad (19)$$

where

$$\Delta_{\text{proxy}} = \mathbb{E}_{(i,j) \sim \mathcal{P}} [\ell(P_{ij}, y_{ij}^*) - \hat{\ell}_{ij}] \quad (20)$$

This proves the decomposition in equation (13).

A.2 Proof of Lemma 3.1

Let $\bar{y}_{ij} = \frac{1}{K} \sum_{k=1}^K y_{ij}^{(k)}$ denote the averaged pxtr-derived soft label. For the binary pairwise ranking loss used in equation (4), we have

$$\ell(p, y) = -y \log p - (1 - y) \log(1 - p) \quad (21)$$

Since this loss is affine with respect to y , the averaged proxy loss can be rewritten as

$$\hat{\ell}_{ij} = \frac{1}{K} \sum_{k=1}^K \ell(P_{ij}, y_{ij}^{(k)}) = \ell(P_{ij}, \bar{y}_{ij}) \quad (22)$$

Therefore, the sample-level proxy bias is

$$\begin{aligned} \Delta_{ij} &= \ell(P_{ij}, y_{ij}^*) - \ell(P_{ij}, \bar{y}_{ij}) \\ &= -\left(y_{ij}^* - \bar{y}_{ij}\right) \log P_{ij} + \left(y_{ij}^* - \bar{y}_{ij}\right) \log(1 - P_{ij}) \\ &= \left(y_{ij}^* - \bar{y}_{ij}\right) \log \frac{1 - P_{ij}}{P_{ij}} \end{aligned} \quad (23)$$

Taking the absolute value gives

$$|\Delta_{ij}| = \left| \log \frac{1 - P_{ij}}{P_{ij}} \right| \cdot \left| y_{ij}^* - \frac{1}{K} \sum_{k=1}^K y_{ij}^{(k)} \right| \quad (24)$$

Under Assumption 1, $P_{ij} \in [\epsilon, 1 - \epsilon]$, and thus

$$\left| \log \frac{1 - P_{ij}}{P_{ij}} \right| \leq \log \frac{1 - \epsilon}{\epsilon} = L \quad (25)$$

Hence, $|\Delta_{ij}| \leq L \cdot \left| y_{ij}^* - \frac{1}{K} \sum_{k=1}^K y_{ij}^{(k)} \right|$, which completes the proof.

A.3 Proof Sketch of Proposition 3.2

Let C_{ij} denote the level of pxtr conflict for item pair (v_i, v_j) . A larger C_{ij} indicates stronger disagreement among the K pxtr ranking pairwise labels $\{y_{ij}^{(k)}\}_{k=1}^K$.

From Lemma 3.1, the sample-level proxy bias satisfies

$$|\Delta_{ij}| \leq L \cdot \left| y_{ij}^* - \frac{1}{K} \sum_{k=1}^K y_{ij}^{(k)} \right| \quad (26)$$

From the equality form in Lemma 3.1, we have

$$|\Delta_{ij}| = \left| \log \frac{1 - P_{ij}}{P_{ij}} \right| \cdot |y_{ij}^* - \bar{y}_{ij}| \quad (27)$$

Since $P_{ij} \in [\epsilon, 1 - \epsilon]$, the coefficient $\left| \log \frac{1 - P_{ij}}{P_{ij}} \right|$ is finite and non-negative. Therefore, the conditional expectation of $|\Delta_{ij}|$ with respect to a conflict level C_{ij} can be written as

$$\mathbb{E}[|\Delta_{ij}| \mid C_{ij}] = \mathbb{E} \left[\left| \log \frac{1 - P_{ij}}{P_{ij}} \right| \cdot |y_{ij}^* - \bar{y}_{ij}| \mid C_{ij} \right] \quad (28)$$

For item pairs with larger pxtr conflict, the averaged proxy label $\bar{y}_{ij}$ has a larger expected deviation from the true satisfaction label. Formally, define

$$g_m(c) = \mathbb{E}[|y_{ij}^* - \bar{y}_{ij}| \mid C_{ij} = c] \quad (29)$$

We use the following monotonic characterization to formalize the relation between pxtr conflict and the unobservable proxy-label mismatch:

$$g_m(c_1) \leq g_m(c_2), \quad \forall c_1 \leq c_2 \quad (30)$$

Combining this condition with the equality form of Lemma 3.1, whose coefficient is non-negative and bounded, we define

$$g_\Delta(c) = \mathbb{E}[|\Delta_{ij}| \mid C_{ij} = c], \quad (31)$$

This supports the following monotonic tendency:

$$g_\Delta(c_1) \leq g_\Delta(c_2), \quad \forall c_1 \leq c_2 \quad (32)$$

On the other hand, Assumption 2 gives the monotonic relation between pxtr conflict and pairwise uncertainty. Define

$$g_U(c) = \mathbb{E}[U_{ij} \mid C_{ij} = c]. \quad (33)$$

Then Assumption 2 implies

$$g_U(c_1) \leq g_U(c_2), \quad \forall c_1 \leq c_2 \quad (34)$$

Thus, both $g_\Delta(c)$ and $g_U(c)$ are non-decreasing functions of the same conflict variable C_{ij} . This shows that the expected proxy-bias magnitude and the expected pairwise uncertainty vary in the same direction as pxtr conflict increases. Therefore, $|\Delta_{ij}|$ and U_{ij} exhibit a positive expectation-level association through their common dependence on pxtr conflict, which establishes the expectation-level association in Proposition 3.2.

B Implementation and Deployment Details

Algorithm 1 summarizes the training and online serving procedures of UAME. During training, UAME predicts both the satisfaction score μ_i and the predictive uncertainty σ_i^2 for each candidate item. The uncertainty is used to construct the probabilistic pairwise ranking loss and the uncertainty-aware adaptive weights. During online serving, only μ_i is used as the final ranking score, while σ_i^2 is not involved in ranking-score computation. Therefore, UAME does not introduce additional ranking-stage inference latency compared with the corresponding backbone model. In practice, the online serving path remains the same as the original end-to-end ensemble ranking backbone, except that the trained μ_i branch is used for final scoring. This makes UAME compatible with existing industrial serving systems without changing the online ranking pipeline.

For numerical stability, the uncertainty branch outputs the log-variance $\log \sigma_i^2$ instead of directly predicting σ_i^2 . The variance is

Algorithm 1 Training and Serving of UAME

Require: Candidate set $\mathcal{I}_{cand}(u, c)$, user/context features, item features, pxtr scores $\{e_i^{(k)}\}_{k=1}^K$, hyperparameters α, β, γ

Ensure: Trained UAME model and online ranking list

```

1: //Offline Training Stage
2: for each training batch do
3:   Encode user, context, item features, and pxtr scores with the
   backbone ranking network.
4:   Predict  $\mu_i$  and  $\log \sigma_i^2$  for each candidate item  $v_i$ .
5:   Obtain  $\sigma_i^2$  from  $\log \sigma_i^2$ .
6:   for each valid item pair  $(v_i, v_j) \in \mathcal{D}$  do
7:     Construct pxtr ranking pairwise labels  $y_{ij}^{(k)} = \mathbb{I}\{e_i^{(k)} > e_j^{(k)}\}$ .
8:     Compute  $P_{ij} = \Phi\left((\mu_i - \mu_j) / \sqrt{\sigma_i^2 + \sigma_j^2}\right)$ .
9:     Compute pairwise uncertainty  $U_{ij} = \sigma_i^2 + \sigma_j^2$ .
10:  end for
11:  Normalize  $\{U_{ij}\}$  and compute  $\omega_{ij} = \gamma \cdot \frac{U_{ij} - \min(U)}{\max(U) - \min(U) + \epsilon}$ .
12:  Compute  $\mathcal{L}_{WPPR}$  by equation (4).
13:  Compute  $\mathcal{L}_{reg}$  by equation (5).
14:  Compute  $\mathcal{L}_{aux}$  by equation (6).
15:  Optimize  $\mathcal{L}_{final} = \sum_{k=1}^K \mathcal{L}_{WPPR}^{(k)} + \alpha \mathcal{L}_{reg} + \beta \mathcal{L}_{aux}$ .
16: end for
17: //Online Serving Stage
18: for each online request do
19:   Encode candidate items with the trained backbone ranking
   network.
20:   Predict the satisfaction score  $\mu_i$  for each candidate item.
21:   Rank candidate items by  $\mu_i$  and return the final ranked list.
22: end for

```

To further analyze the relationship between satisfaction label bias and learned uncertainty in Section 4.5.2, we provide a bucket-level visualization in Figure 4. We compute the satisfaction label bias for each item pair as $B_{ij} = \left| y_{ij}^{q*} - \frac{1}{K} \sum_{k=1}^K y_{ij}^{(k)} \right|$, where y_{ij}^{q*} is the questionnaire-based pairwise satisfaction label.

As shown in Figure 4, item pairs are grouped into five buckets according to B_{ij} , from very low to very high satisfaction label bias. The gray bars denote the number of samples in each bucket, and the blue curve reports the mean pairwise uncertainty within each bucket. The uncertainty increases monotonically as the satisfaction label bias becomes larger, providing an intuitive visualization of the positive relationship between label bias and uncertainty. This trend is consistent with the Pearson and Spearman correlations reported in Section 4.5.2.

then obtained by exponentiation to ensure non-negativity. In addition, a small constant ϵ is added to the denominator of uncertainty normalization and to probability-related computations to avoid division by zero, overflow, or undefined logarithms.

C Additional Analysis of Satisfaction Label Bias and Uncertainty

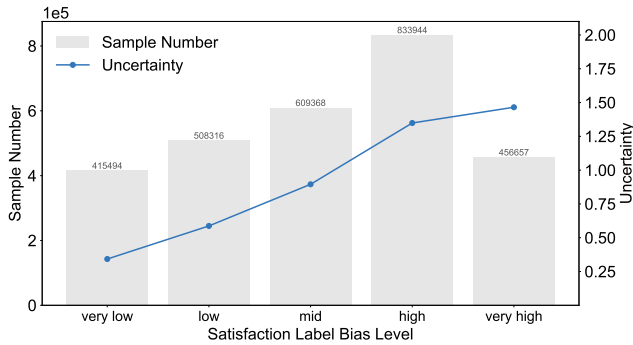


Figure 4: Bucket-level visualization of satisfaction label bias and learned uncertainty.