arXiv:2508.05093v2 [cs.IR] 3 Sep 2025

An End-to-End Multi-objective Ensemble Ranking Framework for Video Recommendation

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

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

Runtong Li*
Kuaishou Technology
Beijing, China
lirt15@tsinghua.org.cn

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

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

Zixiu Wang*
Kuaishou Technology
Beijing, China
wangzixiu@kuaishou.com

Lantao Hu*
Kuaishou Technology
Beijing, China
hulantao@kuaishou.com

Han Li*
Kuaishou Technology
Beijing, China
lihan08@kuaishou.com

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

ABSTRACT

We propose a novel **E**nd-to-end **M**ulti-objective **E**nsemble **R**anking framework (EMER) for the multi-objective ensemble ranking module, which is the most critical component of the short video recommendation system. EMER enhances personalization by replacing manually-designed heuristic formulas with an end-to-end modeling paradigm. EMER introduces a meticulously designed loss function to address the fundamental challenge of defining effective supervision for ensemble ranking, where no single ground-truth signal can fully capture user satisfaction. Moreover, EMER introduces novel sample organization method and transformer-based network architecture to capture the comparative relationships among candidates, which are critical for effective ranking. Additionally, we have proposed an offline-online consistent evaluation system to enhance the efficiency of offline model optimization, which is an established yet persistent challenge within the multi-objective ranking domain in industry. Abundant empirical tests are conducted on a real industrial dataset, and the results well demonstrate the effectiveness of our proposed framework. In addition, our framework has been deployed in the primary scenarios of Kuaishou, a short video recommendation platform with hundreds of millions of daily active users, achieving a 1.39% increase in overall App Stay Time and a 0.196% increase in 7-day user Lifetime(LT7), which are substantial improvements.

CCS CONCEPTS

• **Recommender systems**;

KEYWORDS

Recommender system, Multi-objective, Ensemble Ranking

ACM Reference Format:
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. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

INTRODUCTION

Short video sharing platforms attract a substantial number of users by providing valuable content. When a user accesses a video platform, her satisfaction may manifest through diverse behaviors, which can be clustered into two groups. One group are user explicit interactions, such as liking, sharing, commenting, etc. The other group are user implicit feedback, such as watch time, long viewing, etc. The ranking module is the most critical component of the recommendation system, which is ultimately responsible for determining which videos are presented to the user.

The paradigm for ranking modules widely adopted in industry is a two-stage process. The first stage employs large-scale models to estimate user satisfaction in multiple dimensions. The second stage consolidates these multiple objectives into a single scalar value used for final ranking. In industry, scalarization is predominantly achieved through manually designed heuristic formulas. Although these formulas offer advantages such as low implementation cost, flexible adjustment, and strong interpretability, they exhibit significant limitations in personalization and modeling complex non-linear interactions.

To enhance the personalization and nonlinear modeling capbility of multi-objective ensemble ranking modules, the industry has attempted to upgrade them to an end-to-end modeling paradigm. IntEL [11] explored a model-based approach; however, it retained reliance on a linear ranking formula, with the formula weights being learned by the model. A fundamental limitation of IntEL is the inherently

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 ACM 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, July 2017, Washington, DC, USA
© 2025 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

* These authors contributed equally to this work.

面向视频推荐的端到端多目标融合排序框架

arXiv:2508.05093v2 [cs.IR] 3 Sep 2025

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

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

Runtong Li*
Kuaishou Technology
Beijing, China
lirt15@tsinghua.org.cn

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

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

Zixiu Wang*
Kuaishou Technology
Beijing, China
wangzixiu@kuaishou.com

Lantao Hu*
Kuaishou Technology
Beijing, China
hulantao@kuaishou.com

Han Li*
Kuaishou Technology
Beijing, China
lihan08@kuaishou.com

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

摘要

我们针对多目标融合排序模块提出了一种新颖的端到端多目标融合排序框架（EMER）；该模块是短视频推荐系统最关键的组成部分。EMER 用端到端建模范式取代人工设计的启发式公式，从而增强个性化能力。由于不存在能够完整刻画用户满意度的单一真实信号，如何为融合排序定义有效的监督信号是一项根本性挑战；EMER 引入了精心设计的损失函数来应对这一挑战。此外，EMER 引入了新的样本组织方法和基于Transformer 的网络架构，以捕捉候选项之间的比较关系，这些关系对于有效排序至关重要。我们还提出了一套离线与在线一致的评估体系，以提高离线模型优化的效率；这在工业界多目标排序领域是一项由来已久且持续存在的挑战。我们在真实工业数据集上开展了大量实证测试，结果充分证明了所提框架的有效性。此外，我们的框架已部署于快手的主要场景。快手是一个拥有数亿日活跃用户的短视频推荐平台，部署后整体App 停留时长提升了 1.39%，7 日用户生命周期（LT7）提升了 0.196%，取得了显著改进。

关键词

推荐系统，多目标，融合排序

ACM 引用格式： Tiantian He*、Minzhi Xie*、Runtong Li*、Xiaoxiao Xu*、Jiaqi Yu*、Zixiu Wang*、Lantao Hu*、Han Li* 和 Kun Gai*。2025。面向视频推荐的端到端多目标融合排序框架。收录于《ACM 会议论文集》（Conference'17）。ACM，美国纽约州纽约市，8 页。https://doi.org/10.1145/nnnnnnn.nnnnnnn

引言

短视频分享平台通过提供有价值的内容吸引了大量用户。当用户访问视频平台时，其满意度可能通过多种行为体现出来，这些行为可分为两类。一类是用户的显式互动，例如点赞、分享、评论等。另一类是用户的隐式反馈，例如观看时长、长时间观看等。排序模块是推荐系统最关键的组成部分，最终负责决定向用户展示哪些视频。

工业界广泛采用的排序模块范式包含两个阶段。第一阶段使用大规模模型从多个维度估计用户满意度。第二阶段将这些目标整合为用于最终排序的单个标量值。在工业界，标量化主要通过人工设计的启发式公式实现。尽管这些公式具有实现成本低、调整灵活、可解释性强等优点，但在个性化以及复杂非线性交互建模方面存在显著局限。

为增强多目标融合排序模块的个性化和非线性建模能力，工业界尝试将其升级为端到端建模范式。IntEL [11] 探索了一种基于模型的方法；然而，它仍然依赖线性排序公式，由模型学习公式的权重。IntEL 的一项根本局限在于，线性公式本身的

CCS 概念

• 推荐系统；

允许出于个人或课堂用途免费制作本作品全部或部分内容的数字副本或纸质副本，前提是副本不得用于盈利或商业获益目的的制作或分发，且副本首页须载明本声明和完整引文。本作品中由 ACM 以外的其他方持有版权的组成部分，其版权必须得到尊重。允许在注明出处的情况下进行摘要摘录。其他形式的复制、再版、在服务器上发布或向邮件列表再分发，均须事先取得特别许可和 / 或支付费用。请向 permissions@acm.org 申请许可。Conference'17，2017 年 7 月，美国华盛顿特区。© 2025 Association for Computing Machinery。ACM
ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00 https://doi.org/10.1145/nnnnnnn.nnnnnnn

* These authors contributed equally to this work.

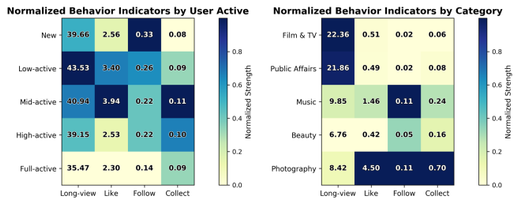


Figure 1: Heatmaps of normalized user behavior indicators. (left) User behaviors normalized by activity level, lower-activity users tend to express satisfaction through follows, with new users exhibiting approximately 135% higher (0.33 vs 0.14) follow rates than fully active users. (right) User behaviors normalized by content category, users demonstrate a higher likelihood of expressing satisfaction via "longer viewing" for "film/TV/entertainment" content, while preferring "like, follow and collect" for music content.

constrained nonlinear capability of the linear formula. UREM subsequently sought to augment IntEL by introducing an unsupervised loss function, aiming to maximize the model's perceptual capacity regarding input information. Nevertheless, existing research lacks a comprehensive discussion of the challenges related to practical deployment.

The core challenges in developing an end-to-end multi-objective ranking model are threefold: 1) Defining effective supervision signals: As illustrated in Figure 1, user satisfaction manifests through heterogeneous behavioral patterns that vary significantly across both user profiles and content taxonomies. That is to say, there is no single ground-truth signal can fully capture user satisfaction. In other words, unlike click-through rate prediction task, there is no unified ground-truth label for multi-objective ensemble ranking task. 2) Ensuring consistency between offline metrics and online performance: Objectives such as watch time and user interactions are susceptible to confounding factors. In our practical experience of optimizing user experience, we consistently observe a discrepancy: Offline model evaluations show improvements in AUC metrics for both watch time and interactions, yet online deployment yields divergent outcomes: increased watch time accompanied by a decline in interaction metrics. 3) Modeling comparative preferences among candidate items: Traditional recommendation system models estimate a user's satisfaction level with a given individual video. However, in real-world recommendation scenarios, the focus shifts to the ranking among multiple candidate videos, that is, to the comparative relationships between the videos within the candidate set.

In this paper, we propose an end-to-end multi-objective ensemble ranking framework to address the above three core challenges. To address the challenge of designing loss function, we introduce the concept of "Relative Advantage Satisfaction", a personalized, comparative measure of user preference. We implement this metric via a pairwise ranking loss. Furthermore, to improve ranking performance across multiple objectives, we incorporate a multi-dimensional AUC-based auxiliary loss. During training, an Advantage Evaluator dynamically adjusts the weights of this AUC loss in real-time, constraining the model towards "self-evolution" in the direction of comprehensively superior AUC performance. To address the issue of inconsistency between offline and online evaluations for watch time and user interactions, we propose a novel interaction metric: Interaction Probability per Unit Time (**IPUT**). Through leveraging the backdoor criterion from causal inference, the **IPUT** metric eliminates the confounding factor of watch time inherent in interactions. Furthermore, we designed a sophisticated sample organization scheme that aggregates all candidate videos to rank within a single request into one consolidated sample. A Transformer-based model architecture is specifically designed to process this aggregated set of candidate videos per request as its input, thereby enabling the model to capture the comparative relationships among the candidate videos. We summarize the main contributions as follows:

- As far as we know, we are the first to clarify that the core challenges in model-based personalized multi-objective ranking are threefold: 1) the inherent difficulty in defining supervision signals, 2) establishing an offline-online consistent evaluation system, and 3) modeling comparative preferences among candidate items.
- To address the three core challenges in model-based personalized multi-objective ranking, we propose a comprehensive framework that includes the design of the loss function, the sample organization scheme, the model architecture, and the evaluation methodology.
- We verify the effectiveness of our proposed framework through extensive experiments conducted on a large scale industrial dataset, and successfully apply it in a real-world large scale recommender system bringing a considerable performance improvement.

RELATED WORK

Short Video Recommendation

User-generated content platforms featuring short videos have increasingly become an integral part of daily life. Recently, increasing research has focused on optimizing short video recommendation systems to enhance the accuracy and efficiency of recommending user-preferred content. Deep sequential models, especially GRU [6] and Transformer architectures [19], have been popular for modeling user behaviors. They enhance interest modeling by capturing patterns in users' historical interactions [9, 16, 24, 25]. Subsequently, research has shifted toward integrating multi-behavioral data to infer satisfaction more accurately. MIND [10] retrieves multiple interest vector representations of users for retrieval, while ComiRec [4] optimizes for diversity. Additionally, other research examines user interests from perspectives such as session analysis [12] and multi-scale time [17]. To jointly capture signals from diverse user feedback and optimize for multiple objectives,multi-task learning frameworks such as MMOE [13] and PLE [18] have been proposed. An alternative approach to better understand user preferences is to collect explicit feedback through surveys [5]. In recent developments, researchers have begun incorporating long-term interest [15, 16, 20, 21] into these models to reflect the evolution of user interests over time. In this paper, we evaluates "relative user satisfaction" by examining various behaviors and their dynamic changes. By continuously refining this measure, we aim to better approximate users' true satisfaction levels and improve long-term personalization.

Conference'17, July 2017, Washington, DC, USA Tiantian He*, Minzhi Xie*, Runtong Li*, Xiaoxiao Xu*, Jiaqi Yu*, Zixiu Wang*, Lantao Hu*, Han Li*, and Kun Gai*

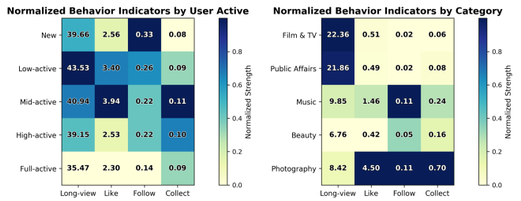


Figure 1: Heatmaps of normalized user behavior indicators. (left) User behaviors normalized by activity level, lower-activity users tend to express satisfaction through follows, with new users exhibiting approximately 135% higher (0.33 vs 0.14) follow rates than fully active users. (right) User behaviors normalized by content category, users demonstrate a higher likelihood of expressing satisfaction via "longer viewing" for "film/TV/entertainment" content, while preferring "like, follow and collect" for music content.

非线性表达能力受到固有限制。随后，UREM 尝试通过引入无监督损失函数来增强 IntEL，旨在最大化模型对输入信息的感知能力。然而，现有研究缺乏对实际部署相关挑战的全面讨论。

开发端到端多目标排序模型的核心挑战有三个方面：1）定义有效的监督信号：如图 1 所示，用户满意度通过异质的行为模式体现出来，这些模式在不同用户画像和内容类别之间存在显著差异。也就是说，不存在能够完整刻画用户满意度的单一真实信号。换言之，与点击率预测任务不同，多目标融合排序任务没有统一的真实标签。2）确保离线指标与在线表现一致：观看时长和用户互动等目标容易受到混杂因素的影响。在优化用户体验的实践中，我们不断观察到一种不一致现象：离线模型评估显示，观看时长和互动两方面的 AUC 指标均有提升，但在线部署却产生了不同的结果，即观看时长增加的同时互动指标下降。3）对候选项之间的比较偏好进行建模：传统推荐系统模型估计用户对某个单独视频的满意程度。然而，在真实推荐场景中，关注点转向多个候选视频之间的排序，即候选集合内视频之间的比较关系。

本文提出了一种端到端多目标融合排序框架，以应对上述三个核心挑战。针对损失函数设计的挑战，我们引入了"相对优势满意度"的概念，这是一种衡量用户偏好的个性化比较指标。我们通过成对排序损失实现这一指标。此外，为提升多个目标上的排序表现，我们引入了基于多维 AUC 的辅助损失。在训练过程中，优势评估器实时动态调整该 AUC 损失的权重，约束模型朝着 AUC 表现全面占优的方向"自我进化"。针对观看时长和用户互动的离线与在线评估不一致问题，我们提出了一种新的互动指标：单位时间互动概率（IPUT）。利用因果推断中的后门准则，IPUT 指标消除了互动中固有的观看时长这一混杂因素。此外，我们设计了精细的样本组织方案，将一次请求内所有待排序的候选视频聚合为一个统一样本。我们专门设计了基于 Transformer 的模型架构，以每次请求聚合后的候选视频集合为输入，使模型能够捕捉候选视频之间的比较关系。我们的主要贡献总结如下：

- 据我们所知，我们首次明确指出，基于模型的个性化多目标排序面临三项核心挑战：1）监督信号定义本身的困难，2）建立离线与在线一致的评估体系，以及 3）对候选物品之间的比较偏好进行建模。
- 为解决基于模型的个性化多目标排序中的这三项核心挑战，我们提出了一个完整框架，涵盖损失函数设计、样本组织方案、模型架构和评估方法。
- 我们在大规模工业数据集上开展了大量实验，验证了所提框架的有效性，并将其成功应用于真实的大规模推荐系统，带来了显著的性能提升。

相关工作

短视频推荐

以短视频为特色的用户生成内容平台已日益成为日常生活不可或缺的一部分。近年来，越来越多的研究致力于优化短视频推荐系统，以提高推荐用户偏好内容的准确性和效率。深度序列模型，尤其是 GRU [6] 和 Transformer 架构 [19]，已被广泛用于用户行为建模。它们通过捕捉用户历史交互中的模式来增强兴趣建模 [9, 16, 24, 25]。随后，研究转向整合多种行为数据，以更准确地推断满意度。MIND [10] 获取用户的多个兴趣向量表示用于召回，而 ComiRec [4] 则针对多样性进行优化。此外，其他研究从会话分析 [12] 和多尺度时间 [17] 等视角考察用户兴趣。为联合捕捉不同用户反馈中的信号并优化多个目标，研究者提出了 MMOE [13] 和 PLE [18] 等多任务学习框架。另一种更好地理解用户偏好的方法是通过问卷调查收集显式反馈 [5]。近期，研究者开始将长期兴趣 [15, 16, 20, 21] 纳入这些模型，以反映用户兴趣随时间的演变。本文通过考察各种行为及其动态变化来评估"相对用户满意度"。通过持续完善这一度量，我们希望更好地逼近用户的真实满意度水平，并改进长期个性化效果。

Ranking Ensemble

Modern recommendation systems often deploy multiple ranking models simultaneously to capture different aspects of user preferences. To combine the outputs into a unified result, ranking ensemble techniques have been widely explored. These techniques aim to leverage the complementary strengths of individual models to improve overall ranking performance. Ranking ensemble methods are generally divided into unsupervised and supervised categories. Unsupervised methods follow predefined rules to merge rankings without requiring annotated data. Typical examples include Median Rank Aggregation [8]. In industrial systems, it is common practice to design weighted fusion formulas that integrate multiple ranking objectives into a unified score. These formulas typically have fixed structures, while the parameters such as objective weights are often manually set based on heuristics or prior knowledge. Some studies have explored search-based strategies like grid search and the Cross-Entropy Method to identify more effective configurations under offline evaluation metrics. Recently, growing efforts have been made to develop and optimize supervised rank ensemble methods. Particularly, model-based techniques, such as the IntEL [11], have demonstrated strong potential in learning effective rank aggregation strategies. Other approaches—such as those based on reinforcement learning [23], Differential Evolution [1, 2], and genetic programming methods like Evolutionary Rank Aggregation [14]—have also been investigated. Collectively, these approaches highlight the increasing diversity of algorithmic strategies developed to enhance ranking ensemble performance in modern recommendation systems.

PROPOSED METHOD

In this section, we provide a comprehensive overview of EMER. Firstly, we describe the details of EMER framework, including model architecture and loss function design, which is illustrated in Figure 2. Afterwards, we introduce an offline-online consistent evaluation system that enhances the efficiency of model optimization.

EMER framework

Sample Construction & Feature Encoding. The essence of ranking lies in determining candidates' relative positions through mutual comparison. EMER highlights the comparative relationship information via the designated feature engineering method and the transformer-based model architecture, based on the introduction of a specific sample organization scheme.

Sample Organization: When organizing the training data, we focus on two key goals: 1) Overcoming exposure bias inherent in learning solely from observed items. 2) Facilitating straightforward item comparison for the model. To achieve these, we meticulously organize training samples at the individual user request level, which means that all candidate items of a request, including exposed and unexposed items, are grouped together to form a consolidated training sample as shown in Figure 2.

Feature Engineering: To empower EMER's ability to model comparative relationships, we design features to capture each item's relative position within the candidate set. NormalizedRanks for different Pxtrs are introduced to provide explicit positional information, helping the model understand how the items stack up against the others in the given set, where

$$\text{NormalizedRanks} = \frac{\text{original rank}}{\text{total number of candidate items}}$$

Model Architecture. EMER employs a transformer-based network to explicitly capture the complex relationships among candidates. This enables the model to weigh each item's influence on others, thereby better understanding their relative comparisons within the set. In such way, the final predicted score reflects both intrinsic item quality and its relative standing within the candidate set.

Learning Objectives & Optimization Strategy

Defining effective loss function is a significant challenge, as user satisfaction is pretty difficult to quantify uniformly. EMER introduces a label named "Relative Advantage Satisfaction" based on the comparable post-exposure feedback within a recommendation request, and presents a solution from a multi-objective optimization perspective to concurrently enhance the model's ranking capability across multiple dimensions of user satisfaction. A self-evolving training scheme is introduced to ensure non-degraded ranking capability across all dimensions.

Modeling Relative Advantage Satisfaction Labels. Relative advantage satisfaction is defined based on posterior feedback (such as likes, follows, watch time et al.) observed after user receiving recommendations, which often serve as the key evidences for evaluating user satisfaction. Given the inherently personalized and varied nature of user behavior, quantifying an absolute score from user feedback is exceptionally difficult. EMER addresses this challenge by defining satisfaction via relative comparisons, identifying the "better" option based on user feedback.

A crucial principle is the hierarchy of "Many Positives" over "Single Positive" over "No Positive": 1) "Many Positives" refers to instances where a user interacts with a recommended item in multiple positive ways. For example, a user might not only *like* an item (one Positive) but also *share* it. 2) "Single Positive" occurs when a user provides one distinct type of positive feedback for a recommended item. This could be a *like* or a *share* on a video. 3) "No Positive" signify that the user provides no positive feedback on an item or even explicitly dislikes it.

For an individual user, relative satisfaction increases with the quantity of positive feedback received, and decreases as the number of negative feedback grows. Following the principle, pairwise logistic loss is employed to encourage pred $\hat{y}_i$ to be significantly higher than $\hat{y}_j$ for all positive pairs D:

$$\begin{aligned} L_{\text{posterior}} &= -\sum_{(i,j)\in D} \log(P(x_i \triangleright x_j)) \\ &= -\sum_{(i,j)\in D} \log(\text{sigmoid}(\hat{y}_i - \hat{y}_j)) \end{aligned} \tag{1}$$

Although posterior feedback offers valuable insight into user satisfaction, it suffers from two intrinsic drawbacks: exposure bias (arising from models learning solely from shown items) and sparse positive signals (due to the infrequent nature of explicit user feedback). To mitigate these limitations, we introduce a complementary solution from a multi-objective optimization perspective.

排序融合

现代推荐系统通常同时部署多个排序模型，以捕捉用户偏好的不同方面。为将这些输出合并为统一结果，排序融合技术得到了广泛研究。这些技术旨在利用各个模型的互补优势，提升整体排序性能。排序融合方法通常分为无监督和有监督两类。无监督方法遵循预定义规则合并排序，无需标注数据。典型例子包括中位数排序聚合 [8]。在工业系统中，常见做法是设计加权融合公式，将多个排序目标整合为一个统一分数。这些公式通常具有固定结构，而目标权重等参数往往依据启发式方法或先验知识手动设置。一些研究探索了网格搜索和交叉熵方法等基于搜索的策略，以找到在离线评估指标下更有效的配置。近年来，开发和优化有监督排序融合方法的工作日益增多。尤其是 IntEL [11] 等基于模型的技术，在学习有效的排序聚合策略方面表现出很强的潜力。其他方法也得到了研究，例如基于强化学习 [23]、差分进化 [1, 2] 的方法，以及进化排序聚合 [14] 等遗传编程方法。总体而言，这些方法表明，为提升现代推荐系统中的排序融合性能而开发的算法策略正日益多样化。

所提方法

本节全面介绍 EMER。首先，我们描述 EMER 框架的细节，包括模型架构和损失函数设计，如图 2 所示。随后，我们介绍一种离线与在线一致的评估体系，以提高模型优化的效率。

EMER 框架

样本构建与特征编码。排序的本质在于通过相互比较来确定候选项的相对位置。EMER 在引入特定样本组织方案的基础上，通过专门设计的特征工程方法和基于 Transformer 的模型架构，突出比较关系信息。

样本组织：在组织训练数据时，我们关注两个关键目标：1）克服仅从已观测物品中学习所固有的曝光偏差。2）使模型能够直接比较物品。为实现这些目标，我们在单次用户请求层面精心组织训练样本，即将一次请求中的所有候选物品，包括已曝光和未曝光物品，聚合为一个整合后的训练样本，如图 2 所示。

特征工程：为增强 EMER 对比较关系的建模能力，我们设计了特征来捕捉每个物品在候选集中的相对位置。我们引入不同 Pxtr 的 NormalizedRanks，以提供显式的位置信息，帮助模型理解在给定集合中各物品与其他物品相比处于何种位置，其中

$$\text{NormalizedRanks} = \frac{\text{original rank}}{\text{total number of candidate items}}$$

模型架构。EMER 采用基于 Transformer 的网络，显式捕捉候选项之间的复杂关系。这使模型能够权衡每个物品对其他物品的影响，从而更好地理解它们在集合中的相对比较关系。通过这种方式，最终预测分数既反映物品的内在质量，也反映其在候选集中的相对位置。

学习目标与优化策略

定义有效的损失函数是一项重大挑战，因为用户满意度很难进行统一量化。EMER 基于一次推荐请求内可比较的曝光后反馈，引入了名为"相对优势满意度"的标签，并从多目标优化的视角提出解决方案，以同时提升模型在用户满意度多个维度上的排序能力。我们引入了一种自演进训练方案，以确保所有维度上的排序能力均不退化。

相对优势满意度标签建模。相对优势满意度根据用户接收推荐后观测到的后验反馈（如点赞、关注、观看时长等）定义，这些反馈通常是评估用户满意度的关键依据。由于用户行为本质上具有个性化和多样性，从用户反馈中量化出绝对分数极为困难。EMER 通过相对比较来定义满意度，根据用户反馈识别"更好"的选项，从而应对这一挑战。

一个关键原则是："多个正反馈"优于"单个正反馈"，而"单个正反馈"又优于"无正反馈"：1）"多个正反馈"是指用户以多种正向方式与推荐条目互动的情形。例如，用户不仅可能对某个条目进行 *like*（一个正反馈），还可能对其进行 *share*。2）"单个正反馈"是指用户对某个推荐条目提供一种明确的正反馈。例如，对某个视频进行一次 *like*或一次 *share*。3）"无正反馈"表示用户没有对某个条目提供任何正反馈，甚至明确表示不喜欢。

对于单个用户，相对满意度随正反馈数量的增加而提高，并随负反馈数量的增加而降低。遵循这一原则，我们采用成对逻辑损失，促使 pred $\hat{y}_i$显著高于$\hat{y}_j$，这一要求适用于所有正样本对 D:

$$\begin{aligned} L_{\text{posterior}} &= -\sum_{(i,j)\in D} \log(P(x_i \triangleright x_j)) \\ &= -\sum_{(i,j)\in D} \log(\text{sigmoid}(\hat{y}_i - \hat{y}_j)) \end{aligned} \tag{1}$$

尽管后验反馈为了解用户满意度提供了有价值的信息，但它存在两个固有缺陷：曝光偏差（模型仅从已展示条目中学习所致）和正向信号稀疏（用户很少提供显式反馈所致）。为缓解这些局限，我们从多目标优化的角度引入一种互补方案。

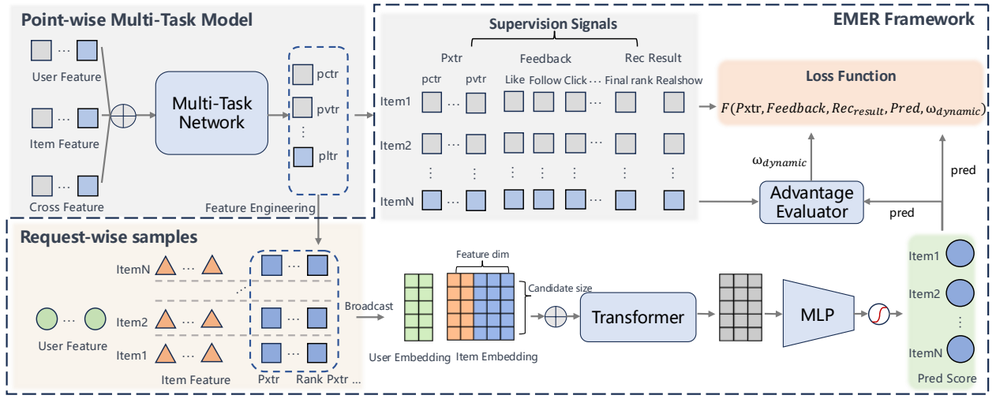


Figure 2: An illustration of the EMER framework.

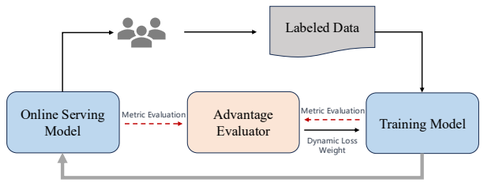


Figure 3: End-to-end pipeline with online self-evolution.

Enhancing multi-dimension ranking capability. In large-scale recommendation systems, multi-task model (Figure 2) provides rich prior signals ($Pxtrs$). The availability of these diverse signals for every candidate is crucial because it enables the model to effectively compare all candidates. Integrating these prior indicators into our user satisfaction definition framework allows for a more robust and forward-looking estimate of satisfaction.

However, effectively synthesizing and optimizing across such a conflicting set of objectives presents a significant challenge for most multi-objective ensemble ranking framework. Instead of combining all prior signals into one label, EMER treats each prior signal as a distinct supervision label, aiming to optimize ranking performance for each. This is predicated on the assumption that each signal reflects a unique facet of user satisfaction. Consequently, improving individual prior signal performance directly improves corresponding satisfaction facet. Therefore, systemic improvements across all prior signals collectively lead to heightened overall user satisfaction.

To achieve improvements across all prior signals, we need to find: 1)A proper loss function with direct alignment with ultimate ranking goal; 2)An effective way to deal with conflicts in multi-objective learning.

To align the learning objective with the ranking goal, EMER targets Area Under Curve (AUC), a standard metric for ranking quality in recommendation. However, direct optimization of AUC is intractable due to its non-differentiability and non-decomposable nature. To address this, we adopt a differentiable pairwise surrogate loss [3, 7]. This formulation approximates AUC maximization by encouraging the predicted score of a more preferred item to be higher than a less preferred one. The label is designed as:

$$y_{pxtr_{\text{ij}}} = \begin{cases} 1 & \text{if } pxtr_i \triangleright pxtr_j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $pxtr$ represents one of $Pxtrs$. Positive sample is labeled 1 for item pair when $item_i$ has higher $pxtr$ than $item_j$ for each user.

The differentiable, pairwise surrogate loss for AUC encourages the probability $P_{i,j}$ that the predicted score $\hat{y}_i$ to be significantly higher than $\hat{y}_j$ for all positive samples.

$$P_{i,j} = \text{sigmoid}(\hat{y}_i - \hat{y}_j) \tag{3}$$

$$L_{pxtr} = -\sum_{(i,j)\in D} y_{pxtr_{\text{ij}}} \log(P_{i,j}) + (1 - y_{pxtr_{\text{ij}}}) \log(1 - P_{i,j}) \tag{4}$$

where L_{pxtr} is the loss for a single indicator $pxtr$. The total loss for all $Pxtrs$ is calculated as:

$$L_{prior} = -\frac{1}{N} \sum_{pxtr}^{Pxtrs} w_{pxtr} \cdot L_{pxtr} \tag{5}$$

where N denotes the number of $Pxtrs$, w_{pxtr} is the weighting coefficients for different objectives (discussed in Self-Evolving Optimization Scheme).

Notably, prior-based satisfaction indicators offer two key advantages as supervision signals. First, they provide denser, continuous values—unlike sparse, binary feedback—which allows the model to learn more nuanced satisfaction levels. Second, they are unconstrained by exposure, eliminating the inherent exposure bias of observed feedback by providing a more comprehensive and unbiased signal for all items.

By combining prior and posterior satisfaction signals, we achieve a more complete and robust definition of user satisfaction. The final training loss is:

$$Loss = L_{posterior} + L_{prior} \tag{6}$$

Conference'17, July 2017, Washington, DC, USA Tiantian He*, Minzhi Xie*, Runtong Li*, Xiaoxiao Xu*, Jiaqi Yu*, Zixiu Wang*, Lantao Hu*, Han Li*, and Kun Gai*

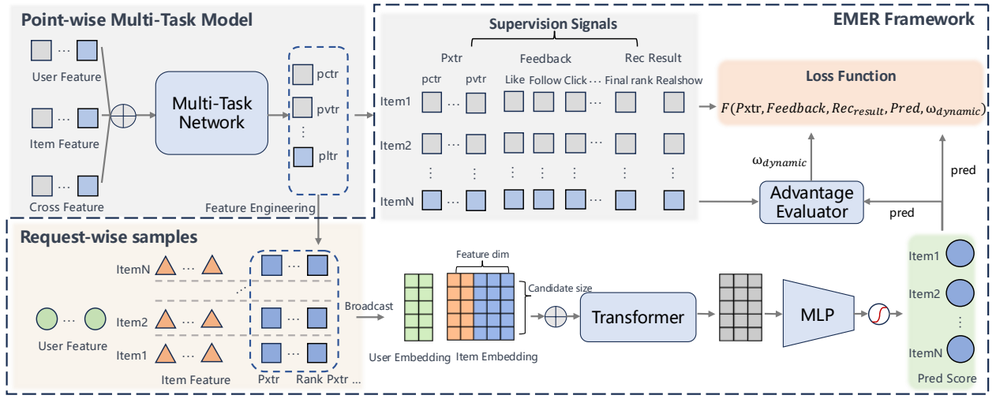


Figure 2: An illustration of the EMER framework.

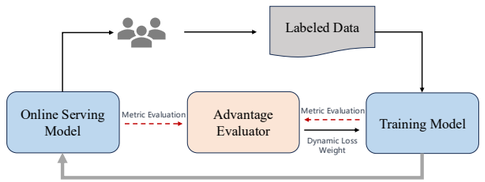


Figure 3: End-to-end pipeline with online self-evolution.

增强多维度排序能力。在大规模推荐系统中，多任务模型（图2）提供了丰富的先验信号($Pxtrs$）。每个候选条目都具备这些多样化信号，这一点至关重要，因为它使模型能够有效比较所有候选条目。将这些先验指标融入我们的用户满意度定义框架，可以对满意度作出更稳健且更具前瞻性的估计。

然而，如何有效综合这一组相互冲突的目标并进行优化，对大多数多目标融合排序框架而言都是一项重大挑战。EMER 将每个先验信号视为独立的监督标签，旨在优化各个信号对应的排序表现，而不是将所有先验信号合并为一个标签。这一做法基于如下假设：每个信号都反映了用户满意度的一个独特方面。因此，提升单个先验信号的表现，可以直接改善相应方面的满意度。由此，对所有先验信号进行系统性改进，便能共同提升整体用户满意度。

要改善所有先验信号的表现，我们需要找到：1）与最终排序目标直接一致的适当损失函数；2）处理多目标学习中冲突的有效方法。

为使学习目标与排序目标一致，EMER 以曲线下面积（AUC）为优化目标；AUC 是推荐中衡量排序质量的标准指标。然而，由于 AUC 不可微且不可分解，直接优化它十分困难。为此，我们采用一种可微的成对替代

损失 [3, 7]。这一形式通过促使更受偏好条目的预测分数高于较少受偏好的条目，来近似实现 AUC 最大化。标签设计如下：

$$y_{pxtr_{\text{ij}}} = \begin{cases} 1 & \text{if } pxtr_i \triangleright pxtr_j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $pxtr$ represents one of $Pxtrs$. Positive sample is labeled 1 for item pair when $item_i$ has higher $pxtr$ than $item_j$ for each user.

The differentiable, pairwise surrogate loss for AUC encourages the probability $P_{i,j}$ that the predicted score $\hat{y}_i$ to be significantly higher than $\hat{y}_j$ for all positive samples.

$$P_{i,j} = \text{sigmoid}(\hat{y}_i - \hat{y}_j) \tag{3}$$

$$L_{pxtr} = -\sum_{(i,j)\in D} y_{pxtr_{\text{ij}}} \log(P_{i,j}) + (1 - y_{pxtr_{\text{ij}}}) \log(1 - P_{i,j}) \tag{4}$$

其中， L_{pxtr}为单个指标 $pxtr$的损失。所有 $Pxtrs$的总损失计算如下：

$$L_{prior} = -\frac{1}{N} \sum_{pxtr}^{Pxtrs} w_{pxtr} \cdot L_{pxtr} \tag{5}$$

其中， N表示 $Pxtrs$的数量， w_{pxtr}是不同目标的加权系数（将在“自演化优化方案”中讨论）。

值得注意的是，基于先验的满意度指标作为监督信号具有两个关键优势。首先，相较于稀疏的二元反馈，它们提供了更密集的连续值，使模型能够学习更细致的满意度水平。其次，它们不受曝光限制，能够为所有条目提供更全面、无偏的信号，从而消除观测反馈固有的曝光偏差。

通过结合先验和后验满意度信号，我们得到了更完整、更稳健的用户满意度定义。最终的训练损失为：

$$Loss = L_{posterior} + L_{prior} \tag{6}$$

Self-Evolving Optimization Scheme. Approximating towards optimal solution for multi-objective is crucial, as different objectives often exhibit inherent conflicts. In EMER, we propose a "Self-Evolving" optimization scheme, employing an "Advantage Evaluator" to personally and dynamically compute the weights of the loss (w_{pxtr} in Equation 5), as shown in Figure 3.

Self-Evolving: EMER introduces a novel concept of model self-evolving, which goes beyond static model updates to enable continuous and adaptive improvement. Unlike traditional approaches where models are retrained periodically on new data or with fixed optimization objectives, EMER is designed to "learn to learn" and dynamically adjust its internal mechanisms based on real-time performance and evolving user behaviors.

Consider a standard supervised learning scenario where a model $f(x;\Theta)$ aims to map an input x to a output $\hat{y}$. The model's performance is quantified by a loss function $L(f(x;\Theta), y)$ over a dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$ with label y. The optimization process seeks to minimize the loss by iteratively updating the parameters Θ:

$$\Theta_{t+1} = \Theta_t - \eta \nabla_\Theta \mathcal{L}(\Theta_t) \tag{7}$$

where η is the learning rate, $\mathcal{L}(\Theta)$ represents the multi-objective loss function (equivalent to L_{prior} in Equation 5) with w_{pxtr} as weighting coefficient to balance the contributions of different objectives.

Unlike methods relying on static weighting coefficients, Self-Evolving scheme employs completely personalized, dynamic and adaptive weighting coefficients. This is achieved by incorporating the performance of both present model $f(x;\Theta_t)$ and previous model $f(x;\Theta_{t-1})$ into the current weight calculation.

$$w_{pxtr} = \mathrm{AE}(f(x;\Theta_t); f(x;\Theta_{t-1})) \tag{8}$$

where w_{pxtr} is strictly positive and AE is the Advantage Evaluator defined in the paragraph below. Notably, EMER does not require to save old model version to get $f(x;\Theta_{t-1})$. We exploit the fact that the current serving model reflects Θ_{t-1} due to online continual training, while the latest model with updated gradients represents Θ_t.

Advantage Evaluator: As mentioned before, Advantage evaluator doesn't measure absolute performance. It quantifies the *relative advantage* of current model compared to historical baselines.

$$\mathrm{AE}(\Theta_t;\Theta_{t-1}) = \frac{Metric(f(x_i;\Theta_{t-1}), y_i)}{Metric(f(x_i;\Theta_t), y_i)} \tag{9}$$

where $Metric(.)$ is the quantitative measure used to evaluate the ranking performance. Choosing one or multiple proper evaluation metrics is also crucial. Different tasks and business goals demand different ways of measurements. Focusing on the highly relevant items ranking performance, we try some different metrics (discussed in Ablation Studies).

Our self-evolving scheme provides a highly efficient and robust multi-objective optimization process. It intelligently eliminates the need for manual parameter tuning by dynamically and adaptively weighting each objective based on its performance. This fine-grained tuning operates at a per-request level, allowing the model to continuously adapt to personalized user preferences and shifting behaviors. The time-based self-comparison further enhances robustness by ensuring fair and accurate weight computation regardless of the inherent scale differences between objectives.

Evaluation methodology

The offline-online inconsistency problem is a tough challenge within the multi-objective ranking domain in industry. This issue arises when a model that performs exceptionally well on offline evaluation metrics fails to achieve similar performance gains when deployed in online environment. This happened in our experiment: models consistently showed better AUC for both watch time and interactions in offline evaluations. But once deployed, we noticed watch time went up, while interactions actually dropped.

This gap highlights a fundamental issue we call the Decoupling Paradox which is deeply rooted in the confounding factor of watch time inherent in interactions.

Specifically, models are typically trained offline to optimize the cumulative interaction probability ($pxtr$) for individual items, aiming to maximize the likelihood of a user interacting with each separate item. However, in the online environment, the true goal is to maximize total cumulative interactions within a user's entire session. To achieve this, the model needs to increase the "probability density" of interactions, striving to generate as many effective interactions per unit of time as possible.

Consider a practical example: Assume that a user totally spends 20 seconds in a session and there are three candidates. Each is represented by (id, p_watchtime, p_interation): (1, 20s, 0.8), (2, 10s, 0.7), (3, 5s, 0.4). The pxtr-optimal sort for interaction ($0.8 > 0.7 > 0.4$) yields 0.8 expected interactions, while the probability density sort for interaction ($0.4/5s > 0.7/10s > 0.8/20s$) delivers at least 1.1 expected interactions. This example highlights that the offline pxtr is not a direct proxy for the "probability density" required online. This fundamental misalignment in objectives means that even if a model performs exceptionally well on offline pxtr, it cannot guarantee a proportional increase in online total interactions.

To eliminate the confounding factor of watch time inherent in interactions, we reformulate the prior interaction $pxtr$ to novel Interaction Probability per Unit Time ($IPUT$)

$$pxtr_{IPUT} = \frac{pxtr}{p_watchtime} \tag{10}$$

These new indicators shift the focus from merely an interaction likelihood to quantifying how efficiently that interaction occurs within a given time budget. This offers significantly more consistent performance between offline evaluations and online deployment.

EXPERIMENT

In this section, we compare the performance of EMER framework to previous ensemble methods. First, we introduce the basic settings of our ranking model. Then, we demonstrate the advantages of our method through both the performance of an online A/B test and the analysis of intermediate results. Finally, we highlight the necessity of each module in EMER through an ablation analysis.

Experimental Setup

Our model is employed in the ranking stage of two primary scenarios on the short video platform. The training dataset consists of recommendation logs generated while users engage with the short video application. Each sample represents a recommendation interaction, including a (user, candidate videos) pair, along with the corresponding feedback. The model is trained incrementally, processing over

自演化优化方案。由于不同目标之间往往存在固有冲突，逐步逼近多目标最优解至关重要。在 EMER 中，我们提出了一种"自演化"优化方案，利用"优势评估器"以个性化、动态的方式计算损失权重（w_{pxtr}，见公式 5），如图 3所示。

自演化：EMER 引入了一种全新的模型自演化概念，超越静态模型更新，实现持续且自适应的改进。传统方法定期利用新数据重新训练模型，或采用固定的优化目标；与之不同，EMER 旨在"学会学习"，并根据实时表现和不断变化的用户行为，动态调整其内部机制。

考虑一个标准的监督学习场景：模型$f(x;\Theta)$旨在将输入 x映射为输出 $\hat{u}$。模型的表现通过损失函数 $L(f(x;\Theta), y)$在数据集 $D = \{(x_i, y_i)\}_{i=1}^{N}$上进行量化，该数据集的标签为 u。优化过程通过迭代更新参数 Θ来最小化损失：

$$\Theta_{t+1} = \Theta_t - \eta \nabla_\Theta \mathcal{L}(\Theta_t) \tag{7}$$

其中， η为学习率， $\mathcal{L}(\Theta)$表示多目标损失函数（等价于 L_{prior} ，见公式 5），并以w_{pxtr}作为加权系数来平衡不同目标的贡献。

与依赖静态加权系数的方法不同，自演化方案采用完全个性化、动态且自适应的加权系数。具体而言，该方案将当前模型 $f(x;\Theta_t)$和先前模型$f(x;\Theta_{t-1})$的表现共同纳入当前权重的计算。

$$w_{pxtr} = \mathrm{AE}(f(x;\Theta_t); f(x;\Theta_{t-1})) \tag{8}$$

其中， w_{pxtr}严格为正， AE为下文定义的优势评估器。值得注意的是，EMER 无需保存旧版本模型来获取 $f(x;\Theta_{t-1})$。我们利用了以下事实：

由于采用在线持续训练，当前提供服务的模型反映了 Θ_{t-1}，而梯度更新后的最新模型则代表 Θ_t。

优势评估器：如前所述，优势评估器并不衡量绝对表现，而是量化当前模型相对于历史基线的相对优势。

$$\mathrm{AE}(\Theta_t;\Theta_{t-1}) = \frac{Metric(f(x_i;\Theta_{t-1}), y_i)}{Metric(f(x_i;\Theta_t), y_i)} \tag{9}$$

其中， $Metric(.)$是用于评估排序表现的量化度量。选择一个或多个适当的评估指标同样至关重要。不同的任务和业务目标需要不同的衡量方式。我们聚焦于高度相关条目的排序表现，尝试了几种不同的指标（将在"消融实验"中讨论）。

我们的自演化方案提供了高效且稳健的多目标优化过程。它根据各个目标的表现，动态、自适应地赋予权重，从而智能地免去了手动调参的需要。这种精细调整在单次请求层面进行，使模型能够持续适应个性化的用户偏好和不断变化的行为。基于时间的自我比较进一步增强了稳健性，确保不论目标之间存在怎样的固有尺度差异，都能公平、准确地计算权重。

评估方法

在工业界的多目标排序领域，离线与在线表现不一致是一个棘手的挑战。该问题表现为：一个在离线评估指标上表现出色的模型，部署到在线环境后却无法取得相似的性能收益。我们的实验中就出现了这种情况：在离线评估中，模型在观看时长和互动两方面的 AUC 都持续更优。然而，部署之后，我们发现观看时长增加了，互动量却实际下降了。

这一差距揭示了一个我们称为"解耦悖论"的根本问题，其深层根源在于互动中固有的观看时长这一混杂因素。

具体而言，模型通常在线下训练，以优化单个物品的累计互动概率（$pxtr$），目标是最大化用户与每个独立物品发生互动的可能性。然而，在在线环境中，真正的目标是最大化用户整个会话内的累计互动总量。为实现这一目标，模型需要提高互动的"概率密度"，力求在单位时间内产生尽可能多的有效互动。

考虑一个实际例子：假设用户在一次会话中总共花费 20 秒，且有三个候选项。每个候选项由标识符、预测观看时长和预测互动概率表示：候选项 1 的预测观看时长为 20 秒，预测互动概率为 0.8；候选项 2 的预测观看时长为 10 秒，预测互动概率为 0.7；候选项 3 的预测观看时长为 5 秒，预测互动概率为 0.4 。针对互动的 pxtr 最优排序（$0.8 > 0.7 > 0.4$）产生的期望互动次数为 0.8，而按互动概率密度排序（$0.4/5s > 0.7/10s > 0.8/20s$）产生的期望互动次数至少为 1.1 。这个例子说明，离线 pxtr 并不能直接替代在线所需的"概率密度"。这种目标上的根本错位意味着，即使模型在离线 pxtr 上表现极佳，也无法保证在线互动总量获得成比例的增长。

为消除互动中固有的观看时长这一混杂因素，我们将原有的互动指标 $pxtr$重新定义为新的单位时间互动概率（$IPUT$）

$$pxtr_{IPUT} = \frac{pxtr}{p_watchtime} \tag{10}$$

这些新指标将关注点从单纯的互动可能性，转向量化在给定时间预算内发生互动的效率。这使离线评估与在线部署之间的表现具有显著更高的一致性。

实验

本节将 EMER 框架的性能与以往的融合方法进行比较。首先，我们介绍排序模型的基本设置。随后，我们通过在线 A/B 测试的表现和中间结果分析，展示所提方法的优势。最后，我们通过消融分析说明 EMER 中各模块的必要性。

实验设置

我们的模型应用于短视频平台两个主要场景的排序阶段。训练数据集由用户使用短视频应用时产生的推荐日志构成。每个样本代表一次推荐交互，包含一个（用户，候选视频）对及相应的反馈。模型采用增量训练方式，每日处理超过

ten billion samples per day, with each sample containing approximately 500 candidate videos. The training is performed using the Adam optimizer with an initial learning rate of 5×10^{-6}. We evaluate the model's performance against two baseline methods:

- Fusion formula (FF). The current online ensemble method of the recommendation platform. The ranking is determined by the score of each item, calculated using the formula $score = f(\theta_1, \texttt{pvtr})\times f(\theta_2, \texttt{pctr})\times\cdots$, where f is a simple smooth function and $\{\theta_1, \theta_2, \cdots\}$ represents the parameters of the function. Both the function and its parameters are manually configured according to the online performance.
- UREM [22]. An ensemble model that combines ranking similarity and reconstruction loss in an unsupervised manner to compute the final ranking score for each item.

Overall Performance

We evaluate the methods based on their performance in online A/B tests. A range of user satisfaction metrics are encompassed as follows:

- retention metric: 7-day Lifetime (LT7).
- implicit feedback metrics: APP stay time, watch time and video view.
- explicit feedback metrics: actions such as like, follow, comment, forward, etc.

Table 1: A/B Performance Comparison between EMER and the Baseline Method. The values in the table represent the improvement of EMER over the baseline method in the online A/B test.

Metrics	Scenario#1 EMER vs FF	Scenario#2 EMER vs FF
LT7	+0.196%	+0.133%
AppStayTime	+1.392%	+1.199%
WatchTime	+1.542%	+2.283%
VideoView	+1.044%	+2.996%
Like	+0.435%	-0.270%
Follow	+0.212%	+0.381%
Comment	+0.695%	+0.643%
Forward	+0.330%	+1.038%

As shown in Table 1, EMER method significantly outperforms baseline approaches across most metrics. Notably, EMER achieves an increase of over 0.10% in LT7 and more than 1.0% in APP stay time in both scenarios. These are remarkable gains in industrial recommendation systems, demonstrating the effectiveness of EMER in video ranking tasks. The tendencies of LT7 are included in the supplement.

To systematically evaluate these ranking methods, we conduct an analysis of rank consistency between the rankings generated by ensemble methods and those derived from pxtrs. We consider the following various pxtrs that score videos from different perspectives.

- pvtr: predicted watch time of a video.
- pctr/plvtr/pcpr: predicted probability of effectively/long/completely watching a video.
- pltr/pwtr/pcmtr/pftr: predicted probability of liking/following (the author of) /commenting/sharing a video.

To measure the consistency between two rankings, we choose the indicator GAUC $= \frac{\sum_i^{\#\text{user}} \#\text{impression}_{u_i}\cdot \text{AUC}_{u_i}}{\sum_i \#\text{impression}_{u_i}}$, where u_i denotes the i-th user, and $\#\text{impression}_{u_i}$ and AUC_{u_i} represent the corresponding number of impressions and AUC respectively. GAUC is a common applied indicator in recommendation systems. Table 2 indicates that our method has better consistency with most of the input pxtrs.

Table 2: GAUC between Rankings Generated by Ensemble Methods and Pxtrs.

pxtr	Scenario#1			Scenario#2		
	EMER	UREM	FF	EMER	UREM	FF
pvtr	**0.727**	0.663	0.596	0.711	**0.718**	0.601
pctr	**0.689**	0.683	0.596	**0.740**	0.726	0.601
plvtr	**0.704**	0.686	0.612	0.758	**0.761**	0.645
pcpr	0.617	**0.683**	0.579	**0.758**	0.701	0.648
pltr	**0.706**	0.611	0.556	0.648	**0.655**	0.575
pwtr	**0.704**	0.593	0.561	**0.683**	0.629	0.567
pcmtr	**0.706**	0.616	0.567	**0.706**	0.667	0.578
pftr	**0.743**	0.624	0.580	**0.736**	0.687	0.584

Further Analysis

In this section, we further analyze the necessity of each module in EMER through Ablation Studies. Additionally, we briefly explain how our chosen metrics influence video recommendations in the Metric Study.

Ablation Studies. To validate our proposed EMER framework, we conducted comprehensive ablation studies. By systematically removing or replacing key components, we quantify their individual contributions to both offline GAUC and online A/B test performance. Our findings, summarized in Tables 3 and 4, lead to the following conclusions.

Table 3: Offline Ablation Studies in term of GAUC.

Metric	*Fusion Formula*	*EMER-NoComp*	*EMER-NoPost*	*EMER-NoPrior*	*EMER-NoEvolve*	*EMER-NoIPUT*	*EMER*
pctr	0.596	0.668	0.653	0.642	0.688	0.683	**0.689**
pvtr	0.596	0.688	0.684	0.659	**0.734**	0.722	0.727
plvtr	0.612	0.685	0.679	0.661	0.701	0.700	**0.704**
pcpr	0.579	0.602	0.601	0.600	0.551	0.607	**0.617**
pltr	0.556	0.689	0.690	0.693	0.680	0.678	**0.706**
pwtr	0.561	0.697	0.681	0.691	0.684	0.680	**0.704**
pcmtr	0.567	0.681	0.686	0.692	0.619	0.686	**0.706**
pftr	0.580	0.732	0.723	0.721	0.631	0.709	**0.743**

Table 4: Online A/B Test Performance over Fusion Formula.

Metric	*EMER-NoComp*	*EMER-NoPost*	*EMER-NoPrior*	*EMER-NoEvolve*	*EMER-NoIPUT*	*EMER*
AppStayTime	+1.094%	+1.045%	+0.939%	+1.233%	+1.242%	**+1.392%**
WatchTime	+1.309%	+1.253%	+1.013%	**+1.689%**	+1.409%	+1.542%
VideoView	+0.953%	+0.960%	+0.802%	-1.303%	+1.001%	**+1.044%**
Like	-1.740%	-0.539%	-1.688%	-1.494%	-1.649%	**+0.435%**
Follow	-0.482%	-1.525%	+0.045%	-1.393%	-7.948%	**+0.212%**
Comment	-0.720%	-1.178%	-1.162%	-7.414%	-6.664%	**+0.695%**
Forward	-1.235%	-1.448%	-1.542%	-8.088%	-10.826%	**+0.330%**

每天一百亿个样本，每个样本包含约 500 个候选视频。训练采用 Adam 优化器，初始学习率为 5×10^{-6}。我们将模型性能与以下两种基线方法进行比较：

• 融合公式（FF）。推荐平台当前在线使用的融合方法。排序由每个物品的分数决定，该分数通过公式 $score = f(\theta_1, \text{pvtr})\times f(\theta_2, \text{pctr})\times\cdots$ 计算，其中 f 是一个简单的平滑函数，$\{\theta_1, \theta_2, \cdots\}$ 表示该函数的参数。函数及其参数均根据线上表现手动配置。• UREM [22]。一种以无监督方式结合排序相似性与重构损失、计算每个物品最终排序分数的融合模型。

整体性能

我们根据各方法在在线 A/B 测试中的表现进行评估。评估涵盖以下一系列用户满意度指标：

• 留存指标：7日用户生命周期（LT7）。• 隐式反馈指标：App 停留时长、观看时长和视频观看次数。• 显式反馈指标：点赞、关注、评论、转发等行为。

Table 1: A/B Performance Comparison between EMER and the Baseline Method. The values in the table represent the improvement of EMER over the baseline method in the online A/B test.

Metrics	Scenario#1 EMER vs FF	Scenario#2 EMER vs FF
LT7	+0.196%	+0.133%
AppStayTime	+1.392%	+1.199%
WatchTime	+1.542%	+2.283%
VideoView	+1.044%	+2.996%
Like	+0.435%	-0.270%
Follow	+0.212%	+0.381%
Comment	+0.695%	+0.643%
Forward	+0.330%	+1.038%

如表 1 所示，EMER 方法在大多数指标上显著优于基线方法。值得关注的是，在两个场景中，EMER 均使 LT7 提升超过 0.10%，App 停留时长提升超过 1.0%。这些增益在工业推荐系统中相当显著，证明了 EMER 在视频排序任务中的有效性。LT7 的变化趋势见补充材料。

为系统评估这些排序方法，我们分析了融合方法生成的排序与由 pxtr 得到的排序之间的一致性。我们考虑以下从不同视角为视频评分的各类 pxtr。

• pvtr：预测的视频观看时长。• pctr/plvtr/pcpr：预测的有效观看 / 长时间观看 / 完整观看视频的概率。• pltr/pwtr/pcmtr/pftr：预测的点赞视频 / 关注（视频作者）/ 评论视频 / 分享视频的概率。

为衡量两个排序之间的一致性，我们选择

指标 GAUC $= \frac{\sum_i^{\#\text{user}} \#\text{i}}{\sum_i}\#\text{impression}_{u_i}\cdot\text{AUC}_{u_i}\#\text{impression}_{u_i}$，其中 u_i 表示第 i 个用户，$\#\text{impression}_{u_i}$ 和 AUC_{u_i} 分别表示对应的曝光次数和 AUC。GAUC 是推荐系统中常用的指标。表 2 表明，我们的方法与大多数输入 pxtr 具有更好的一致性。

Table 2: GAUC between Rankings Generated by Ensemble Methods and Pxtrs.

pxtr	Scenario#1			Scenario#2		
	EMER	UREM	FF	EMER	UREM	FF
pvtr	**0.727**	0.663	0.596	0.711	**0.718**	0.601
pctr	**0.689**	0.683	0.596	**0.740**	0.726	0.601
plvtr	**0.704**	0.686	0.612	0.758	**0.761**	0.645
pcpr	0.617	**0.683**	0.579	**0.758**	0.701	0.648
pltr	**0.706**	0.611	0.556	0.648	**0.655**	0.575
pwtr	**0.704**	0.593	0.561	**0.683**	0.629	0.567
pcmtr	**0.706**	0.616	0.567	**0.706**	0.667	0.578
pftr	**0.743**	0.624	0.580	**0.736**	0.687	0.584

进一步分析

本节通过消融研究进一步分析 EMER 中各模块的必要性。此外，我们在指标研究中简要说明所选指标如何影响视频推荐。

消融实验。为验证我们提出的 EMER 框架，我们开展了全面的消融实验。通过系统性地移除或替换关键组件，我们量化了各组件对离线 GAUC 和在线 A/B 测试表现的独立贡献。实验结果汇总于表 3和表 4，并得出以下结论。

Table 3: Offline Ablation Studies in term of GAUC.

Metric	*Fusion Formula*	*EMER-NoComp*	*EMER-NoPost*	*EMER-NoPrior*	*EMER-NoEvolve*	*EMER-NoIPUT*	*EMER*
pctr	0.596	0.668	0.653	0.642	0.688	0.683	**0.689**
pvtr	0.596	0.688	0.684	0.659	**0.734**	0.722	0.727
plvtr	0.612	0.685	0.679	0.661	0.701	0.700	**0.704**
pcpr	0.579	0.602	0.601	0.600	0.551	0.607	**0.617**
pltr	0.556	0.689	0.690	0.693	0.680	0.678	**0.706**
pwtr	0.561	0.697	0.681	0.691	0.684	0.680	**0.704**
pcmtr	0.567	0.681	0.686	0.692	0.619	0.686	**0.706**
pftr	0.580	0.732	0.723	0.721	0.631	0.709	**0.743**

Table 4: Online A/B Test Performance over Fusion Formula.

Metric	*EMER-NoComp*	*EMER-NoPost*	*EMER-NoPrior*	*EMER-NoEvolve*	*EMER-NoIPUT*	*EMER*
AppStayTime	+1.094%	+1.045%	+0.939%	+1.233%	+1.242%	**+1.392%**
WatchTime	+1.309%	+1.253%	+1.013%	**+1.689%**	+1.409%	+1.542%
VideoView	+0.953%	+0.960%	+0.802%	-1.303%	+1.001%	**+1.044%**
Like	-1.740%	-0.539%	-1.688%	-1.494%	-1.649%	**+0.435%**
Follow	-0.482%	-1.525%	+0.045%	-1.393%	-7.948%	**+0.212%**
Comment	-0.720%	-1.178%	-1.162%	-7.414%	-6.664%	**+0.695%**
Forward	-1.235%	-1.448%	-1.542%	-8.088%	-10.826%	**+0.330%**

Ablation of Modeling Comparative Relationships: To validate the effectiveness of explicitly modeling candidates' comparative relationships, we created a variant called EMER-NoComp. This model removes the request-level sample grouping, the transformer-based network and *NormalizedRanks* as input. EMER-NoComp scores each item in isolation rather than determining their relative positions. The results show a measurable decline in both offline metrics (Table 3) and online metrics (Table 4). This confirms that modeling comparative relationships within the candidate set is crucial for enhancing item representation and ranking quality.

Ablation of Components in Supervision Signals: EMER's supervision signals have two parts: posterior-based relative satisfaction and prior-based multi-dimensional objectives. We created two variants: EMER-NoPost (removes posterior-based signals) and EMER-NoPrior (removes prior-based signals). Both variants show a performance decline, confirming the individual effectiveness of each component. The decline is more significant with EMER-NoPrior, which validates the crucial role of prior signals due to their density, lack of exposure bias, and capacity to provide multi-dimensional insights.

Ablation of Self-Evolving Optimization Scheme: To investigate the core contribution of the self-evolving scheme, we designed EMER-NoEvolve, a variant that trains using static weights for different objectives. As shown in Table 3 and Table 4, EMER-NoEvolve exhibits a significant performance imbalance across various metrics. For example, online WatchTime increases, but at the cost of significant drops in other metrics (e.g., VideoView -2.347%, Forward -8.418%, and Comment -8.109%). This confirms that removing the self-evolving scheme greatly diminishes the model's ability to resolve multi-objective conflicts. Even with substantial manually weights tuning, EMER consistently outperformed all manually tuned configurations. A deeper analysis (Figure 4 in the Appendix) revealed that EMER's multi-objective loss distribution is more uniform and centralized with lower value than EMER-NoEvolve's. This clearly validates that the self-evolving scheme is crucial for guiding the model's learning direction and ensuring its robustness and stability over time.

Ablation of Interaction Probability per Unit Time: As shown in Table 5, EMER-NoIPUT suffers from a classic decoupling paradox: significant offline improvements in the AUC of $pltr$ and $pwtr$ fail to translate into positive online A/B test results. In contrast, positive offline gains of $pxtr_{IPUT}$ directly align with positive online results. This ablation demonstrates that IPUT module is vital for aligning offline training with real-world online user satisfaction.

Metric Study. Within the Self-Evolving Optimization Scheme, the Advantage Evaluator critically assesses the current model and historical baselines. The choice of its evaluation metric is paramount. We investigated three strategies (Equation 11- 13): $HitRate@K$, $MEAN@K$, and $DCG@K$ shown in the Appendix, using $pxtr$ as a positive signal. As Table 6 shows, $DCG@K$ consistently outperforms $HitRate@K$ and $MEAN@K$ across nearly all offline GAUC metrics. This indicates $DCG@K$, by emphasizing top-ranked item relevance and position, leads to more accurate predictions and better aligns with complex user engagement. Therefore, we adopted the $DCG@K$ as the default evaluation metric for our dynamic weighting.

Table 5: Offline-Online Consistent Evaluation Study.

Metric	*Fusion Formula*	*EMER-NoIPUT*	*EMER*
Offline GAUC Metric			
pltr	0.556	0.678	**0.706**
pwtr	0.561	0.680	**0.704**
pcmtr	0.567	0.686	**0.706**
pftr	0.580	0.709	**0.743**
$pltr_{IPUT}$	0.555	0.516	**0.567**
$pwtr_{IPUT}$	0.556	0.505	**0.587**
$pcmtr_{IPUT}$	0.598	0.521	**0.612**
$pftr_{IPUT}$	0.557	0.512	**0.574**
Online A/B Test over FF			
Like	-	-1.649%	**+0.435%**
Follow	-	-7.948%	**+0.212%**
Comment	-	-6.664%	**+0.695%**
Forward	-	-10.826%	**+0.330%**

Table 6: Offline Metric Study in term of GAUC.

Metric	*HitRate@K*	*MEAN@K*	*DCG@K*(ours)
pctr	0.654	0.674	**0.689**
pvtr	0.699	0.711	**0.727**
plvtr	0.681	0.683	**0.704**
pcpr	0.593	0.584	**0.617**
pltr	0.673	0.676	**0.706**
pwtr	0.689	0.681	**0.704**
pcmtr	0.670	0.688	**0.706**
pftr	0.690	0.721	**0.743**

CONCLUSION

In this paper, we present EMER, a novel end-to-end multi-objective ensemble ranking framework designed to enhance personalization in large-scale recommendation systems. EMER addresses the challenge of defining supervision signals for the ensemble ranking model by employing a meticulously designed loss function. Moreover, EMER introduces novel sample organization method and transformer-based network architecture to capture the comparative relationships among candidates, which are critical for effective ranking. Our framework has been successfully deployed in the main feed of Kuaishou, a short-video recommendation platform that serves hundreds of millions of daily active users, achieving substantial improvements in user engagement and overall system performance. These results demonstrate the practical utility and scalability of EMER in real-world industrial environments.

REFERENCES

[1] Michał Bałchanowski and Urszula Boryczka. 2022. Aggregation of rankings using metaheuristics in recommendation systems. *Electronics* 11, 3 (2022), 369.

[2] Michał Bałchanowski and Urszula Boryczka. 2022. Collaborative rank aggregation in recommendation systems. *Procedia computer science* 207 (2022), 2213–2222.

[3] Toon Calders and Szymon Jaroszewicz. 2007. Efficient AUC optimization for classification. In *European conference on principles of data mining and knowledge discovery*. Springer, 42–53.

[4] Yukuo Cen, Jianwei Zhang, Xu Zou, Chang Zhou, Hongxia Yang, and Jie Tang. 2020. Controllable multi-interest framework for recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery &*

比较关系建模的消融实验：为验证显式建模候选条目间比较关系的有效性，我们构建了名为 EMER-NoComp 的变体。该模型移除了请求级样本分组、基于 Transformer 的网络以及作为输入的 *NormalizedRanks*。EMER-NoComp 独立地为每个条目评分，而不确定它们的相对位置。结果表明，离线指标（表 3）和在线指标（表 4）均出现了可观测的下降。这证实，对候选集合内部的比较关系进行建模，对于增强条目表征和提升排序质量至关重要。

监督信号各组成部分的消融实验：EMER 的监督信号包含两个部分：基于后验的相对满意度，以及基于先验的多维目标。我们构建了两个变体：EMER-NoPost（移除基于后验的信号）和 EMER-NoPrior（移除基于先验的信号）。两个变体的性能均有所下降，证实了各组成部分的独立有效性。其中，EMER-NoPrior 的性能下降更为显著，这验证了先验信号的关键作用，原因在于它们具有密集性、不存在曝光偏差，并且能够提供多维度的信息。

自演化优化方案的消融实验：为探究自演化方案的核心贡献，我们设计了 EMER-NoEvolve 变体，该变体在训练时对不同目标使用静态权重。如表 3和表 4所示，EMER-NoEvolve 在各项指标上的表现明显失衡。例如，在线 WatchTime 上升，却以其他指标大幅下降为代价（如 VideoView -2.347%、Forward -8.418% 和 Comment -8.109%）。这证实，移除自演化方案会大幅削弱模型解决多目标冲突的能力。即便进行了大量手动权重调优，EMER 仍始终优于所有手动调优的配置。进一步分析（附录中的图 4）发现，相较于 EMER-NoEvolve，EMER 的多目标损失分布更均匀、更集中，且取值更低。这清楚地验证了自演化方案对于引导模型学习方向，以及确保模型随时间保持稳健性和稳定性的重要作用。

单位时间互动概率的消融实验：如表 5 所示，EMER-NoIPUT 存在典型的解耦悖论：$pltr$和 $pwtr$的离线 AUC 显著提升，却未能转化为在线 A/B 测试中的正向结果。相比之下，$pxtr_{IPUT}$的离线正向收益与在线正向结果直接一致。该消融实验证明，IPUT 模块对于使离线训练与真实在线用户满意度保持一致至关重要。

指标研究。在自进化优化方案中，优势评估器承担着评估当前模型和历史基线的重要职责，其评估指标的选择至关重要。我们考察了附录中给出的三种策略（公式 11- 13）：$HitRate@K$、$MEAN@K$ 和 $DCG@K$，并使用 $pxtr$作为正向信号。如表 6 所示，在几乎所有离线 GAUC 指标上，$DCG@K$ 均持续优于$HitRate@K$和 $MEAN@K$。这表明，$DCG@K$通过强调排名靠前物品的相关性及其位置，能够给出更准确的预测，并更好地契合复杂的用户参与行为。因此，我们采用 $DCG@K$作为动态加权的默认评估指标。

Table 5: Offline-Online Consistent Evaluation Study.

Metric	*Fusion Formula*	*EMER-NoIPUT*	*EMER*
Offline GAUC Metric			
pltr	0.556	0.678	**0.706**
pwtr	0.561	0.680	**0.704**
pcmtr	0.567	0.686	**0.706**
pftr	0.580	0.709	**0.743**
$pltr_{IPUT}$	0.555	0.516	**0.567**
$pwtr_{IPUT}$	0.556	0.505	**0.587**
$pcmtr_{IPUT}$	0.598	0.521	**0.612**
$pftr_{IPUT}$	0.557	0.512	**0.574**
Online A/B Test over FF			
Like	-	-1.649%	**+0.435%**
Follow	-	-7.948%	**+0.212%**
Comment	-	-6.664%	**+0.695%**
Forward	-	-10.826%	**+0.330%**

Table 6: Offline Metric Study in term of GAUC.

Metric	*HitRate@K*	*MEAN@K*	*DCG@K*(ours)
pctr	0.654	0.674	**0.689**
pvtr	0.699	0.711	**0.727**
plvtr	0.681	0.683	**0.704**
pcpr	0.593	0.584	**0.617**
pltr	0.673	0.676	**0.706**
pwtr	0.689	0.681	**0.704**
pcmtr	0.670	0.688	**0.706**
pftr	0.690	0.721	**0.743**

结论

本文提出了 EMER，这是一种新颖的端到端多目标融合排序框架，旨在增强大规模推荐系统的个性化能力。EMER 使用精心设计的损失函数，解决了融合排序模型监督信号定义的挑战。此外，EMER 引入了新的样本组织方法和基于 Transformer 的网络架构，以捕捉候选项之间的比较关系，这些关系对于有效排序至关重要。我们的框架已成功部署于快手的主信息流；快手是一个服务数亿日活跃用户的短视频推荐平台。部署后，用户参与度和系统整体性能均取得了显著提升。这些结果证明了 EMER 在真实工业环境中的实用性和可扩展性。

REFERENCES

[1] Michał Bałchanowski and Urszula Boryczka. 2022. Aggregation of rankings using metaheuristics in recommendation systems. *Electronics* 11, 3 (2022), 369.

[2] Michał Bałchanowski and Urszula Boryczka. 2022. Collaborative rank aggregation in recommendation systems. *Procedia computer science* 207 (2022), 2213–2222.

[3] Toon Calders and Szymon Jaroszewicz. 2007. Efficient AUC optimization for classification. In *European conference on principles of data mining and knowledge discovery*. Springer, 42–53.

[4] Yukuo Cen, Jianwei Zhang, Xu Zou, Chang Zhou, Hongxia Yang, and Jie Tang. 2020. Controllable multi-interest framework for recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery &*

data mining. 2942–2951.
[5] Konstantina Christakopoulou, Can Xu, Sai Zhang, Sriraj Badam, Trevor Potter, Daniel Li, Hao Wan, Xinyang Yi, Ya Le, Chris Berg, et al. 2022. Reward Shaping for User Satisfaction in a REINFORCE Recommender. *arXiv preprint arXiv:2209.15166* (2022).
[6] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).
[7] Corinna Cortes and Mehryar Mohri. 2003. AUC optimization vs. error rate minimization. *Advances in neural information processing systems* 16 (2003).
[8] Ronald Fagin, Ravi Kumar, and Dandapani Sivakumar. 2003. Efficient similarity search and classification via rank aggregation. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*. 301–312.
[9] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939* (2015).
[10] Chao Li, Zhiyuan Liu, Mengmeng Wu, Yuchi Xu, Huan Zhao, Pipei Huang, Guoliang Kang, Qiwei Chen, Wei Li, and Dik Lun Lee. 2019. Multi-interest network with dynamic routing for recommendation at Tmall. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 2615–2623.
[11] Jiayu Li, Peijie Sun, Zhefan Wang, Weizhi Ma, Yangkun Li, Min Zhang, Zhoutian Feng, and Daiyue Xue. 2023. Intent-aware Ranking Ensemble for Personalized Recommendation *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 1004–1013. https://doi.org/10.1145/3539618.3591702
[12] Fuyu Lv, Taiwei Jin, Changlong Yu, Fei Sun, Quan Lin, Keping Yang, and Wilfred Ng. 2019. SDM: Sequential deep matching model for online large-scale recommender system. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 2635–2643.
[13] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. 2018. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1930–1939.
[14] Samuel Oliveira, Victor Diniz, Anisio Lacerda, and Gisele L Pappa. 2016. Evolutionary rank aggregation for recommender systems. In *2016 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 255–262.
[15] Qi Pi, Weijie Bian, Guorui Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Practice on long sequential user behavior modeling for click-through rate prediction. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2671–2679.
[16] Qi Pi, Guorui Zhou, Yujing Zhang, Zhe Wang, Lejian Ren, Ying Fan, Xiaoqiang Zhu, and Kun Gai. 2020. Search-based user interest modeling with lifelong sequential behavior data for click-through rate prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2685–2692.
[17] Qiaoyu Tan, Jianwei Zhang, Ninghao Liu, Xiao Huang, Hongxia Yang, Jingren Zhou, and Xia Hu. 2021. Dynamic memory based attention network for sequential recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 4384–4392.
[18] Hongyan Tang, Junning Liu, Ming Zhao, and Xudong Gong. 2020. Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. In *Proceedings of the 14th ACM conference on recommender systems*. 269–278.
[19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. arXiv:1706.03762 [cs.CL]
[20] Jing Yan, Liu Jiang, Jianfei Cui, Zhichen Zhao, Xingyan Bin, Feng Zhang, and Zuotao Liu. 2024. Trinity: Syncretizing Multi-/Long-Tail/Long-Term Interests All in One. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 6095–6104.
[21] Congrui Yi, David Zumwalt, Zijian Ni, and Shreya Chakrabarti. 2023. Progressive horizon learning: Adaptive long term optimization for personalized recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 940–946.
[22] Wenhui Yu, Bingqi Liu, Bin Xia, Xiaoxiao Xu, Ying Chen, Yongchang Li, and Lantao Hu. 2024. Unsupervised Ranking Ensemble Model for Recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*, Ricardo Baeza-Yates and Francesco Bonchi (Eds.). ACM, 6181–6189. https://doi.org/10.1145/3637528.3671598
[23] 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.
[24] Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Deep interest evolution network for click-through rate prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 5941–5948.
[25] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1059–1068.

APPENDIX

- *HitRate@K*: The intersection of the top-K items from the pxtr ranking list and the final ranking list.

$$HitRate@K_{pxtr} = \frac{List_{pxtr,K} \cap List_{final,K}}{K} \tag{11}$$

- *MEAN@K*: The mean *pxtr* score of the top-K ranking(r) items from the final ranking list.

$$MEAN@K_{pxtr} = \frac{\sum_{r=1}^{K} pxtr_r}{K} \tag{12}$$

- *DCG@K*: The cumulative gain with a ranking positional (r) discount from the top-K final ranking list.

$$DCG@K_{pxtr} = \sum_{r=1}^{K} \frac{pxtr_r}{\log_2(r+1)} \tag{13}$$

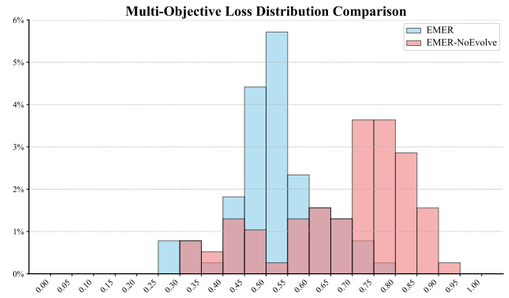


Figure 4: Multi-Objective Loss Distribution Comparison. EMER achieves a lower, more uniform multi-objective loss distribution, while EMER without MSE leads to a skewed distribution with high-loss values.

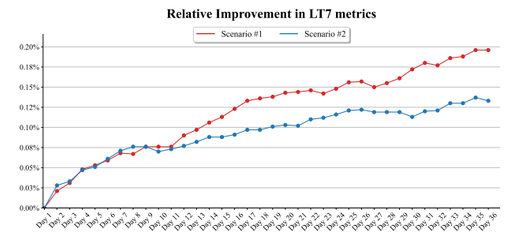


Figure 5: Relative Improvement in LT7 metrics over Fusion Formula.

Conference'17, July 2017, Washington, DC, USA Tiantian He*, Minzhi Xie*, Runtong Li*, Xiaoxiao Xu*, Jiaqi Yu*, Zixiu Wang*, Lantao Hu*, Han Li*, and Kun Gai*

data mining. 2942–2951.
[5] Konstantina Christakopoulou, Can Xu, Sai Zhang, Sriraj Badam, Trevor Potter, Daniel Li, Hao Wan, Xinyang Yi, Ya Le, Chris Berg, et al. 2022. Reward Shaping for User Satisfaction in a REINFORCE Recommender. *arXiv preprint arXiv:2209.15166* (2022).
[6] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).
[7] Corinna Cortes and Mehryar Mohri. 2003. AUC optimization vs. error rate minimization. *Advances in neural information processing systems* 16 (2003).
[8] Ronald Fagin, Ravi Kumar, and Dandapani Sivakumar. 2003. Efficient similarity search and classification via rank aggregation. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*. 301–312.
[9] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939* (2015).
[10] Chao Li, Zhiyuan Liu, Mengmeng Wu, Yuchi Xu, Huan Zhao, Pipei Huang, Guoliang Kang, Qiwei Chen, Wei Li, and Dik Lun Lee. 2019. Multi-interest network with dynamic routing for recommendation at Tmall. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 2615–2623.
[11] Jiayu Li, Peijie Sun, Zhefan Wang, Weizhi Ma, Yangkun Li, Min Zhang, Zhoutian Feng, and Daiyue Xue. 2023. Intent-aware Ranking Ensemble for Personalized Recommendation *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 1004–1013. https://doi.org/10.1145/3539618.3591702
[12] Fuyu Lv, Taiwei Jin, Changlong Yu, Fei Sun, Quan Lin, Keping Yang, and Wilfred Ng. 2019. SDM: Sequential deep matching model for online large-scale recommender system. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 2635–2643.
[13] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. 2018. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1930–1939.
[14] Samuel Oliveira, Victor Diniz, Anisio Lacerda, and Gisele L Pappa. 2016. Evolutionary rank aggregation for recommender systems. In *2016 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 255–262.
[15] Qi Pi, Weijie Bian, Guorui Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Practice on long sequential user behavior modeling for click-through rate prediction. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2671–2679.
[16] Qi Pi, Guorui Zhou, Yujing Zhang, Zhe Wang, Lejian Ren, Ying Fan, Xiaoqiang Zhu, and Kun Gai. 2020. Search-based user interest modeling with lifelong sequential behavior data for click-through rate prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2685–2692.
[17] Qiaoyu Tan, Jianwei Zhang, Ninghao Liu, Xiao Huang, Hongxia Yang, Jingren Zhou, and Xia Hu. 2021. Dynamic memory based attention network for sequential recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 4384–4392.
[18] Hongyan Tang, Junning Liu, Ming Zhao, and Xudong Gong. 2020. Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. In *Proceedings of the 14th ACM conference on recommender systems*. 269–278.
[19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. arXiv:1706.03762 [cs.CL]
[20] Jing Yan, Liu Jiang, Jianfei Cui, Zhichen Zhao, Xingyan Bin, Feng Zhang, and Zuotao Liu. 2024. Trinity: Syncretizing Multi-/Long-Tail/Long-Term Interests All in One. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 6095–6104.
[21] Congrui Yi, David Zumwalt, Zijian Ni, and Shreya Chakrabarti. 2023. Progressive horizon learning: Adaptive long term optimization for personalized recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 940–946.
[22] Wenhui Yu, Bingqi Liu, Bin Xia, Xiaoxiao Xu, Ying Chen, Yongchang Li, and Lantao Hu. 2024. Unsupervised Ranking Ensemble Model for Recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*, Ricardo Baeza-Yates and Francesco Bonchi (Eds.). ACM, 6181–6189. https://doi.org/10.1145/3637528.3671598
[23] 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.
[24] Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Deep interest evolution network for click-through rate prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 5941–5948.
[25] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1059–1068.

附录

- *HitRate@K*：pxtr 排序列表与最终排序列表中前 K 个物品的交集。

$$HitRate@K_{pxtr} = \frac{List_{pxtr,K} \cap List_{final,K}}{K} \tag{11}$$

- *MEAN@K*：平均 *pxtr*得分，计算范围为最终排序列表中排名前 K（r)的物品。

- *DCG@K*：最终排序列表中前 K 个物品经排序位置（r）折扣后的累积增益。

$$DCG@K_{pxtr} = \sum_{r=1}^{K} \frac{pxtr_r}{\log_2(r+1)} \tag{13}$$

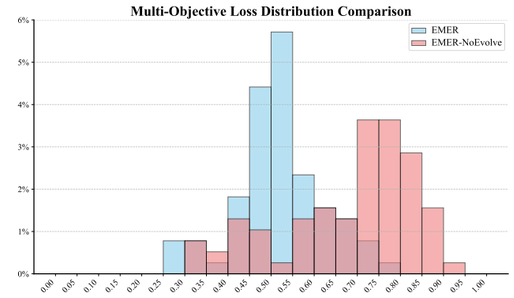


图 **4**：多目标损失分布对比。**EMER** 实现了更低且更均匀的多目标损失分布，而不使用 **MSE** 的 **EMER** 则呈现出损失值较高的偏斜分布。

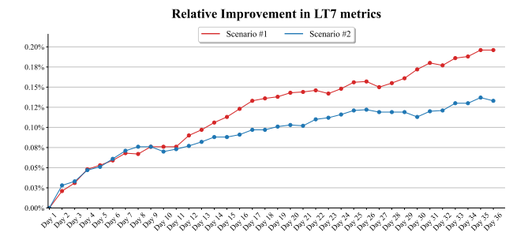


图 **5**：相较于融合公式，**LT7** 指标的相对提升。