

Melody 音乐推荐算法架构设计

Melody: WEB3 音乐革命领导者

Transformer 多头多任务模型

采用统一的 Transformer 编码层提取用户和歌曲的深层特征表示，并在其顶部设置多个任务头。通过共享编码器，一方面利用 Transformer 强大的序列建模能力捕捉用户听歌序列偏好，另一方面让多个任务共同学习用户和内容特征，从而提升泛化性能 researchgate.net。具体包括三个并行的预测头：

- **用户兴趣头**：预测用户对歌曲的偏好程度，可用播放时长、完播率等衡量。模型可输出一个兴趣评分（或完成率的概率）。训练时结合用户的隐式反馈信号（例如实际播放时长占比、跳过/完成标记）和显式反馈（点赞、收藏等）作为监督信号。通过多任务学习将隐式和显式反馈一起建模，可提升对用户真实兴趣的把握 researchgate.net。
- **订阅转化头**：预测用户付费订阅的可能性（如在当前推荐下试用高级服务或订阅的概率）。该头可以根据历史付费行为、试听次数与订阅转化关系等进行训练，目标是提升订阅转化率。
- **推广曝光头**：预测推广歌曲对用户的影响力，例如用户是否会点击或完整播放推广曲，或该推荐对提升歌曲曝光的效果评分。这个头侧重于平台的商业推广目标。

各任务头共享 Transformer 编码器的输出特征表示，但拥有各自的全连接层或注意力层来专门学习该任务相关的信号。如此共享与区分结合的**“共享底

座，多任务塔” **架构，有助于在优化多个目标时共享用户行为模式，同时避免任务间互相干扰 static.aminer.cn。若任务之间存在冲突，还可在多头之间引入例如 **Multi-gate Mixture-of-Experts (MMoE)** 等模块，通过专家网络动态分配不同任务使用的特征子空间，缓解多目标优化的冲突 static.aminer.cn。模型输入包括用户的历史听歌序列、用户画像特征、歌曲内容特征等，由 Transformer 编码层建模序列依赖和高阶特征交互，再输出到各任务头进行多目标预测。

目标函数设计：多反馈信号融合与权重平衡

模型训练采用多任务损失函数，将**用户体验**和**商业目标**相关的损失组合优化。对于用户兴趣任务，可采用回归或分类损失（如预测播放时长占比的回归误差，或完播与否的二分类交叉熵损失）；对于订阅预测和推广效果预测任务，采用适合概率预测的损失（如交叉熵损失）。同时融入显式反馈（如收藏、点赞标记的交叉熵损失）和隐式反馈信号（播放时长、重复收听次数、跳过率等，可转化为偏好分数或分类标签）共同训练 researchgate.net。总损失是各任务损失的加权和：
$$\mathcal{L} = w_1 \mathcal{L}_{\text{兴趣}} + w_2 \mathcal{L}_{\text{订阅}} + w_3 \mathcal{L}_{\text{曝光}}$$
。为了平衡**自然推荐**与**商业推广**，需要谨慎设置损失权重：保证用户兴趣相关指标（播放时长、完播率）的损失权重大于或相当于商业指标，从而确保优化过程首先关注用户体验。在实践中，可以通过**动态调整权重**或**分阶段训练**来平衡：例如先预训练兴趣预测任务，再微调加入订阅和曝光任务，并逐步增加商业目标权重，确保模型不偏离用户体验目标。必要时可在目标函数中增加约束项，惩罚过度牺牲用户体验的策略。例如，当检测到推荐结果的完播率下降超出阈值，可提高兴趣损失权重

或在损失中加入完播率的约束。这样保证模型在优化商业目标时，以不损害用户体验为前提。实践中字节跳动等企业也强调“用户优先”，即宁可牺牲一些短期收入也要保证用户留存体验 tech.huanqiu.com, research.atspotify.com（研究亦发现，如果一个会话中包含过多新奇或推广内容，用户往往会跳过大部分此类歌曲，必须适度控制频率 research.atspotify.com）。因此，损失设计上通过权重平衡和约束，确保**用户体验不受干扰**的同时兼顾商业指标。

个性化推荐策略：动态权重与强化学习调整

不同用户对商业内容容忍度和兴趣侧重点各异，因此对不同用户群体动态调整多任务策略至关重要。系统可以针对用户的历史行为模式和兴趣变化，**自适应地调整多任务损失权重和推荐策略**：例如，对于沉浸式音乐爱好者，可降低推广内容权重、提高个性化兴趣权重；对于有潜在付费意向的用户（如频繁试听高级内容但未订阅），适当提高订阅预测在排序中的权重。实现方法上，可以引入**强化学习或 Bandit 机制**。例如，将推荐策略视为一个 Agent，根据当前用户状态（兴趣分布、近期行为）选择动作（比如推荐列表中商业推广项的比例，或调整权重 w_1, w_2, w_3 ），并根据用户后续反馈获得奖励信号（综合考虑播放时长、完播率、以及付费转化等）。通过强化学习的方法，不断调整策略使长期回报最大化，包括用户留存和商业收益。研究表明，利用强化学习动态调节多任务权重是可行的：比如 WWW 2023 的一项工作提出了强化学习增强的多任务推荐框架 RMTL，使用 Actor-Critic 网络根据情境自动更新各任务损失的权重，从而比固定权重的多任务模型取得更高的指标表现 cnblogs.com。这说明**针对每个用户实时学习最优权重组合**可以有效平衡多目标。除此之外，也可以采用**上下文多臂 Bandit**方法：将自然歌曲推荐和推广

歌曲视为不同臂，根据用户反馈动态调整探索与利用的比例。例如系统可以设定一个 baseline 的用户满意度得分和商业得分，对于每次推荐，根据实时反馈（如用户是否跳过推广歌曲或是否付费）更新对不同策略的信心度，以决定下一轮推荐中各类内容出现的概率 eugeneyan.com, eugeneyan.com。通过 Bandit 算法在探索新策略与利用当前最优策略间取得平衡，逐步收敛到针对该用户的最优策略配置。总之，**个性化多任务权重调整**保证了推荐系统能敏捷响应个体差异，在满足大多数用户体验的同时，针对有商业转化潜力的用户也不过度放弃商业机会。

排序与混排策略：多目标融合的个性化排序

有了多头模型对候选歌曲进行多维度打分，接下来需要设计排序函数将这些分值融合，生成最终的推荐列表。常用策略是对多目标输出进行**加权线性融合**：对于候选歌曲 i ，根据 Transformer 输出的兴趣分数 $S_{\text{兴趣}}(i)$ 、订阅分数 $S_{\text{订阅}}(i)$ 、曝光分数 $S_{\text{曝光}}(i)$ ，计算综合得分：

$$S_{\text{综合}}(i) = \alpha \cdot S_{\text{兴趣}}(i) + \beta \cdot S_{\text{订阅}}(i) + \gamma \cdot S_{\text{曝光}}(i) + \epsilon_i$$
$$S_{\text{综合}}(i) = \alpha \cdot S_{\text{兴趣}}(i) + \beta \cdot S_{\text{订阅}}(i) + \gamma \cdot S_{\text{曝光}}(i) + \epsilon_i$$

其中 α, β, γ 为策略权重， ϵ_i 是微小噪声项防止分数相同或用于探索。权重可以根据用户群体差异设定：例如**新用户**可能更注重培养兴趣粘性， α 权重更高；**老用户**但未订阅的，可以略提高 β 以引导订阅；**已付费用户**则 β 为 0 且适当考虑 γ 提升多样性。除了线性融合，也可训练一个**排序模型 (Ranking Model)**，将多头预测值和其他特征一起作为输入，通过学习式排序 (Learning-to-Rank) 直

接输出综合评分。无论何种融合方式，都需确保推广歌曲不会完全淹没自然喜欢的歌曲。可采用**限制和混排策略**：对列表结果应用业务规则，如每 N 首歌中至少插入 M 首推广歌曲上限，避免连续出现过多广告推荐，以免破坏体验 research.atspotify.com。另外，可针对不同情境下的用户动态调整排序策略，比如用户在通勤场景可能更愿意听熟悉歌曲，此时模型降低新歌探索和商业内容的比重；而在探索新歌的心情时，则可以适度提高新内容曝光。系统可以预先对用户进行分群（如“新用户/资深用户”，“重度音乐发烧友/轻度听众”，“可能付费转化用户/无付费意愿用户”等），**针对群体定制排序策略**：为倾向新内容的用户提高 Discovery 类歌曲权重，为保守型用户则主推熟悉风格歌曲并少插入广告。通过**离线模拟和在线 A/B 测试**不断调整 α, β, γ 在各分群和场景下的取值，实现不同用户看到的列表在满足总体指标的同时更贴合个人偏好。这样的多目标排序融合策略能有效在**相关性（用户兴趣）和多样性/商业收益**之间取得平衡 ijcai.org。

实时优化：近实时响应与反馈闭环

音乐推荐必须跟随用户兴趣的潮起潮落实时演进。为此，系统架构需支持**近实时的数据更新和模型推理**。一方面，在特征层面，将用户最近的行为（如最近几首播放、搜索的歌曲，当前收听时长等）作为**实时特征**输入模型。例如设计短时兴趣模块，将最近几分钟内用户与歌曲的交互序列通过一个小型 Transformer 或者时序模型编码，融合到主模型的决策中，从而使当前推荐结果能够反映用户此刻最新的兴趣。xie.infoq.cn 指出相较按天更新，基于用户最近几秒行为实时调整推荐列表，可使用户当前兴趣“所见即所得”，极大增强用户体验 xie.infoq.cn。因此我们的系统会持续监测用户当前会话中的反

馈：如果用户突然对某一类型歌曲反应热烈（如连续听完多首某风格歌曲），实时提高该风格在候选集中的权重；若用户搜索了某歌曲或歌手，也立刻提升与其相关内容的推荐概率。另一方面，在模型训练层面，引入**反馈闭环**机制：对用户的每次交互（播放完毕、跳过、点赞、订阅等）进行日志采集，经过近实时流处理管道更新用户画像和偏好模型。例如采用增量训练或在线学习方法，持续用最新的数据微调模型的部分参数或更新模型的内部状态。大型互联网应用中常见 **T+0 近线更新**：每隔几分钟将最新一批用户行为合并到训练数据中，更新模型或用户 embedding 向量，从而模型参数每天多次甚至实时刷新 huaweicloud.com, xie.infoq.cn。同时利用**在线 AB 测试**和**反馈评估**，监控推荐效果的核心指标（点击率、完播率、订阅转化等）和用户留存数据，将这些指标的变化作为对模型的实时调控信号：若发现某段时间完播率下降，则可能需要降低推广内容占比或调整推荐列表组成；如果订阅转化低于预期且不影响播放时长，则可以适当在榜单中增添对订阅有帮助的内容并观察效果。通过这种**实时反馈闭环**，模型能够不断自我校正：既快速捕获短期兴趣偏移，又根据长期效果指标调整策略权重，形成持续优化的自适应推荐系统。

综上，该音乐推荐系统架构以 Transformer 为核心，融合多头多任务学习，在训练目标上平衡用户体验与商业目标，并通过个性化的策略权重和实时反馈优化，实现动态、精准的推荐：既让用户持续获得喜爱的歌曲，实现高完播和长播放时长，又在适当时机展示付费或推广内容以达成商业收益，真正做到用户价值和商业价值的双赢。

参考文献：

1. Long et al. *"Multi-task learning for collaborative filtering."*

IJMLC, 2022

[researchgate.net](https://www.researchgate.net)

– 将用户行为划分为显式反馈和隐式反馈，通过多任务模型结合两者提升推荐效果。

2. Ma et al. *"Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts."* **KDD**, 2018

static.aminer.cn

– 提出 MMoE 模块有效解决多任务推荐中不同目标间的冲突，提升多目标预测表现。

3. Abdollahpouri et al. *"Balancing Multi-objective Recommendations."* **RecSys**, 2019

research.atspotify.com

– 强调在音乐推荐中平衡用户满意度和多样性目标的重要性，指出过多新颖内容会导致用户跳过，提高目标平衡策略的必要性。

4. Zhao et al. *"Multi-Task Recommendations with Reinforcement Learning."* **WWW**, 2023

cnblogs.com

- 提出强化学习增强的多任务推荐框架 RMTL，通过 Actor-Critic 动态调整任务权重，在多个公开数据集上优于静态权重的多任务模型。

5. InfoQ 技术社区. *“实时个性化推荐系统”*. 2023

xie.infoq.cn

- 介绍了实时推荐的概念与价值，即根据用户最近行为秒级更新推荐结果，让系统快速响应用户兴趣变化，提高用户体验。