人们为什么跳过音乐？基于深度强化学习的音乐跳过预测

Francesco Meggetto NeuraSearch 实验室，斯特拉斯克莱德大学，英国格拉斯哥
francesco.meggetto@strath.ac.uk

Crawford Revie 斯特拉斯克莱德大学，英国格拉斯哥
cra
wford.revie@strath.ac.uk

John Levine 斯特拉斯克莱德大学，英国格拉斯哥
john.levine@strath.ac.uk

Yashar Moshfeghi NeuraSearch 实验室，斯特拉斯克莱德大学，英国格拉斯哥
yashar.moshfeghi@strath.ac.uk

arXiv:2301.03881v1 [cs.IR] 10 Jan 2023

摘要

音乐推荐系统是我们日常生活中不可或缺的一部分。近年来，研究者在基于黑箱推荐器的方法上投入了大量精力，例如深度强化学习（深度强化学习）。随着这些技术的进步，以及人们对用户数据收集和隐私问题日益增长的担忧，构建负责任的推荐系统也受到广泛关注。成功的音乐推荐系统的一个关键要素，是对用户如何与流媒体内容互动进行建模。通过首先了解这些互动，我们可以获得洞见，从而构建更加透明且负责任的系统。跳过行为就是这类互动的一个例子，它可以作为衡量用户满意度、不满意度或缺乏兴趣的信号。本文研究用户历史数据对于按顺序预测用户跳过行为这一任务的效用。为此，我们将深度强化学习应用于这一分类任务，并对输入状态表示进行事后可解释性分析（SHAP）和消融分析。基于真实世界音乐流媒体数据集（Spotify）的实验结果表明，我们的方法在该任务中优于最先进模型，效果显著。对我们的方法和用户历史数据进行全面分析后，发现该数据集中存在时间数据泄漏问题。研究结果表明，总体而言，在我们提出的深度强化学习模型预测音乐跳过行为时，用户行为特征最具区分力。内容和情境特征的影响较小。这表明，只需收集并利用有限的用户数据，就能预测跳过行为。

CCS 概念

• **信息系统** →**推荐系统**。

关键词

Spotify、音乐、跳过、用户行为、预测、深度强化学习

允许免费复制本作品的部分或全部内容，用于个人或课堂用途，前提是不得为盈利或商业利益而制作或分发副本，并且副本须包含本声明及首页上的完整引文。本作品中第三方组件的版权必须得到尊重。如需用于其他用途，请联系所有者/作者。*CHIIR '23，2023*年*3*月*19*日至*23*日，美国德克萨斯州奥斯汀© 2023 版权所有，归所有者/作者所有。ACM ISBN 979-8-4007-0035-4/23/03。
https://doi.org/10.1145/3576840.3578312

ACM 引用格式：
Francesco Meggetto、Crawford Revie、John Levine 和 Yashar Moshfeghi。2023。人们为什么跳过音乐？基于深度强化学习的音乐跳过预测。发表于*ACM SIGIR*人类信息交互与检索会议（*CHIIR '23*），*2023*年*3*月*19*日至*23*日，美国德克萨斯州奥斯汀。ACM，美国纽约州纽约市，12页。
https://doi.org/10.1145/3576840.3578312

1 引言

近年来，在线音乐流媒体服务（例如 Spotify）取得了显著增长。随着数字音乐发行的发展、此类流媒体服务的成功，以及音乐的普遍可用，一种全新的聆听模式应运而生。用户只需点击几下，就能随时收听任何歌曲。因此，用户的行为以及与这些系统的交互方式发生了显著变化 [19]。音乐推荐系统（MRS）致力于解决如何为用户提供所需支持的问题，帮助他们访问这些庞大的音乐资源库，并找到符合其兴趣和需求的歌曲。近期研究已在基于黑箱的方法上投入了大量精力，例如深度强化学习（DRL） [3, 56]。这是因为用户的行为可能会在切换歌曲时，甚至在同一首歌曲播放过程中的不同时间点，发生剧烈变化。用户行为受到外部因素（潮流趋势）和内部因素（个人兴趣的变化）的影响。用户兴趣和行为不断变化，因此很难学习出能够在任何时刻满足用户特定需求的通用模型；由于需要持续学习和适应，这种情况下就需要采用 DRL [9, 29, 61]。然而，这些进展不可避免地引发了人们对用户数据如何收集、存储和使用的日益关注。这促使研究者对构建负责任的系统和数据收集流程产生了浓厚兴趣 [45]。在考虑收集哪些数据以及将哪些数据提供给模型以衡量用户行为时，应当设定限制。这是因为引入错误和偏差可能带来危害。因此，尽量减少数据特征并筛选高质量的数据特征至关重要。

如今的系统中，显式评分数据相对稀少，因此对隐式反馈进行建模变得愈发重要。例如，在轻松聆听模式下，如自动播放列表或广播流媒体，用户交互会尽量减少。用户一次只会看到一首歌曲。音乐推荐系统几乎完全需要依赖隐式反馈信号（例如跳过或拖动进度，即通过移动光标向前或向后跳转

cursor [35]）来预测满意度和参与度 [29, 61]。通过理解这些交互，可以为构建更加透明且负责任的系统提供见解。跳过是一种信号，可用于衡量用户对平台的满意度、不满或兴趣缺乏，以及参与度 [44]。在轻松聆听模式下，音乐推荐系统通常设计得更加保守，优先利用而非探索，以尽量减少负面反馈（在此情境下即跳过） [53]。因此，其目标之一可能是推荐能够带来最高聆听活动（即未被跳过）的歌曲。然而，用户的跳过行为仍是一个尚未得到充分探索的领域[11, 34, 44]。由于其本质上存在噪声，这是一个颇具挑战性的问题：跳过可能表示一次负面交互，但用户也可能因为最近在别处听过某首喜欢的歌曲而将其跳过。在这项工作中，我们旨在通过全面分析用户历史数据的效用，理解人们为何跳过歌曲。具体而言，我们分析用户行为（例如导致当前播放开始的用户操作）、聆听内容（即所听的歌曲）以及上下文特征（例如一天中的时段）对预测用户音乐跳过行为这一分类任务的影响。我们提出一种新方法，将深度强化学习用于这一分类任务，并对其加以改造，以尽可能贴近基于深度强化学习的音乐推荐系统学习检测跳过行为的方式。

以往对跳过行为的分析揭示了跨歌曲普遍存在的跳过行为；地理位置、音频波动或音乐事件，以及聆听情境信息，都会影响人们如何跳过音乐 [14, 44, 48, 49]。近期，研究者也开始探索深度学习模型在歌曲聆听会话中预测用户序列跳过行为这一任务上的效果 [1, 8, 12, 23, 31, 58, 68]。尽管这些研究为这一方向作出了重要贡献，但其过程通常被视为独立且静态的流程。它们可能无法考虑用户行为的动态特性，也不会直观地针对用户满意度和参与度的长期潜力进行优化 [29, 40, 54, 61, 66, 67]。总的来说，这促使我们探究深度强化学习在预测音乐跳过方面的适用性，并全面考察跳过信号与用户行为、聆听情境及内容之间的关系。本文旨在探讨以下两个重要研究问题：深度强化学习能否应用于用户音乐跳过行为预测任务？如果可以，那么在音乐跳过预测任务中，它是否比深度学习最先进的模型更有效？（**研究问题1**）；哪些历史信息具有判别性，可作为高质量指标，帮助模型预测人们为什么跳过音乐？（**研究问题 2**）。为探究这些研究问题，我们基于真实世界的音乐流媒体数据集（Spotify）开展了广泛研究。全面分析表明，我们的方法行之有效，同时也揭示了历史数据中存在时间性数据泄漏问题。总体而言，我们的研究结果表明，对于所提出的深度强化学习模型而言，预测音乐跳过时最具判别性的特征是一些用户行为特征，而内容特征和情境特征的影响较小。这说明只需有限的用户数据即可预测这种行为，进而为构建新型以用户为中心的音乐推荐系统以及负责任的数据收集流程提供启示。这是全面呈现听众偏好、兴趣和需求的必要一步。本文的主要贡献如下：

- 我们证明了深度强化学习能够适用于根据聆听会话预测用户的跳过行为，并且效果良好。我们设计了一个框架，以拓展深度强化学习在执行此分类任务和离线学习方面的适用性。据我们所知，这是首次将深度强化学习用于这一任务。我们在真实世界的音乐流媒体数据集（Spotify）上，通过实证展示了所提方法的有效性。在平均准确率和首次预测准确率指标方面，我们提出的方法优于最先进模型。

- 我们对所提出的方法进行了全面的事后分析（SHAP）和消融分析，以研究用户的历史数据在检测音乐跳过行为方面的作用。我们揭示了历史数据中存在时间数据泄漏问题。此外，结果表明，在所提出的深度强化学习模型预测音乐跳过行为时，总体而言，用户行为特征最为突出且最具区分性。聆听内容和情境特征的影响较小。

2 相关工作

成功的音乐推荐系统需要随时满足用户的各种需求 [24, 55, 64]。因此，用户建模是一个关键要素。一系列研究试图厘清人格与用户音乐偏好之间的关系 [37, 51, 52]。Volokhin 和 Agichtein [60]提出了音乐聆听意图这一概念，并指出意图有别于情境（用户的活动）。另一个不同且可以说是互补的研究方向，是尝试理解并建模用户与底层平台的交互方式。这是在线流媒体服务长期存在却研究不足的问题 [11]。歌曲之间的跳过行为就是这类交互的一个例子。在音乐聆听会话中对其进行建模和理解，对于理解用户行为至关重要 [44]。跳过行为往往是底层音乐推荐系统唯一可用的信息，因此常被用作推断音乐偏好的代理指标 [53]。

跳过信号已在以往研究中得到应用，例如作为基于启发式的播放列表生成系统中的衡量指标，用于评估 [10, 50],用户满意度 [24, 64], 相关性 [25], , 或作为反事实估计器 [43]。此外，鉴于该信号具有普遍性，且也存在于其他领域，近期研究还考察了它对社交媒体平台广告的影响 [5–7]。尽管数量丰富，但这是一种带有噪声的隐式信号 [53, 62]。跳过某首曲目并不一定意味着用户不喜欢它。关于用户为何跳过歌曲，可以提出多种假设；近期研究表明，人们在跳过歌曲时表现出一种普遍行为，而这种行为受时间、地理位置以及对音频波动或音乐事件的反应所影响 [14, 48, 49]。此外，研究 [57] 表明，通常会听完整首歌曲的人，其聆听时长高于不会听完整首歌曲的人。最近，Meggetto 等人 [44] 提出了一种基于聚类的方法，能够清晰地识别出四种与基于会话的跳过活动相关的用户类型，分别是聆听者、先听后跳过者、先跳过后聆听者和跳过者。这些类型受到聆听会话时长、一天中的时段以及播放列表类型的影响。以往研究的主要局限在于，它们探讨的是聆听情境和内容与跳过行为之间的关系，却没有探讨用户与

平台的交互如何影响跳过行为的检测。这正是本研究要解决的局限。

2019年，Spotify 将音乐跳过预测确定为一项重要挑战，并组织了 序列跳过预测挑战赛 [1]，以探索能够缓解这一问题的方法。该挑战赛侧重于预测聆听会话中的各个曲目是否会被跳过。为应对这一挑战，研究人员提出了多种深度神经网络 [1, 8, 12, 23, 31, 58, 68] 和监督学习 [18] 模型。Afchar 和 Hennequin [2] 提出使用可解释的深度神经网络，通过特征归因来解释跳过行为。尽管神经网络，尤其是循环神经网络（RNN），已被证明能够有效地对序列数据建模，但它们将这一过程视为静态过程。它们并未直观地提供一种机制，以长期优化用户满意度和参与度、实现持续学习，以及对用户行为的动态特性进行建模 [29, 54, 61, 66, 67]。因此，这正是需要深度强化学习（DRL）的情形，而此前尚未对此进行研究或应用。本研究旨在填补这一研究空白。

序列跳过预测挑战赛是一项二元分类任务。尽管迄今受到的关注有限，但研究表明，深度强化学习适用于分类任务且效果显著。它可以帮助分类器学习有利特征 [15, 27]，并从噪声数据中筛选出高质量样本 [17]。Wiering等人 [65] 证明，强化学习确实适用于分类。他们的模型略胜于现有分类器，但训练时间长且需要额外计算资源，这是其主要缺点。随着该领域近期的发展，越来越多的研究显示，基于深度强化学习的方法在分类任务中具有优越性 [17, 27, 28, 39, 42]。具体而言， [27, 28] 中的作者表明，采用原始深度Q网络（深度Q网络）的方法 [47] 优于最先进的算法，且鲁棒性更强。

在这项工作中，我们首次探索深度强化学习在按序预测用户跳过音乐行为这一任务中的适用性。这一研究受到现有方法局限性以及深度强化学习优势的启发。通过全面分析用户历史数据，我们研究了该方法在此任务中的实用性和效果。这项工作是理解人们为何跳过音乐的第一步。

3 方法

本节介绍我们的框架，以便将深度强化学习应用于从聆听会话中按顺序预测用户跳过行为的问题。为此，我们将该问题建模为一个马尔可夫决策过程（MDP），并在强化学习问题的形式化定义中引入一种机制，以正确利用日志交互数据，从而执行离线学习。该框架的详情如下：

状态：它是在离散时间步（即会话中的位置）对聆听会话进行的记录级表示。状态，也就是聆听会话中的一条记录，包含用户关于流的各种上下文信息、用户与平台的交互历史，以及用户所听曲目的相关信息。

[1]https://www.aicrowd.com/challenges/spotify-sequential-skip-prediction-challenge

一个回合是完整的聆听会话，每个会话包含 10 至 20 条记录。

动作：这是一个离散动作空间，用二元指示变量表示相应用户是否会跳过当前曲目。实际上，该问题的形式化定义也可以视为一个二元分类问题 $A = \{0,1\}$，其中 0 表示不执行跳过操作，1 表示跳过。

奖励：若正确预测了跳过分类，则给予 1 的正奖励；否则奖励为 0（即不予惩罚）。

考虑到音乐跳过预测任务的离散动作空间和离策略要求，我们采用深度Q网络（DQN）[2]。这些要求使得我们无法使用深度确定性策略梯度等算法（其动作空间为连续空间），也无法使用近端策略优化（其采用同策略学习）。虽然该问题被建模为马尔可夫决策过程（MDP），但根据定义，它具有部分可观测性（POMDP）。这是因为我们只能获取有关聆听情境和用户的部分信息 [16]。因此，在问题建模中，我们将MDP和POMDP视为等同。这意味着我们不会对状态表示进行任何进一步处理（例如屏蔽某些特征）。

这种分类建模可以视为一场猜测游戏：猜对即可获得正奖励，猜错则不受惩罚。通过折扣因子 γ进行长期优化，可以理解为尽可能准确地猜出一个回合中的多条记录。由于一个回合内的记录（即一次音乐聆听会话）存在序列相关性，因此应采用较高的 γ值。这相当于优化一个回合内猜对的总次数（长期优化），而不是优化即时猜测结果（短期优化）。通过考虑过去的时间点以及与环境的过往交互，深度强化学习智能体能够做出信息充分的决策。

3.1 离线机制

DQN 的标准训练流程完全是在线的。在线学习是一种迭代过程，智能体通过与环境交互来收集新的经验，通常使用其最新学到的策略。随后，这些经验会被用于改进智能体的策略。然而，利用日志数据作为一种（预）训练方式，可能对智能体有所帮助并提供有价值的信息。在离线学习（批量强化学习 [36]）中，智能体的任务则被定义为从静态数据集中学习。策略从日志数据中学习，无需与底层环境交互。虽然我们之前的表述适用于在线学习场景，但在进行离线学习时会带来一个重大问题。错误分类会导致状态转移到一个新状态，但该状态并不属于原始轨迹，因此也未在数据集中表示。智能体会为一条错误生成的轨迹分配并计算（折扣）累积奖励，而这条轨迹与原始轨迹有很大差异。因此，纯离线算法必须完全依赖输入所提供数据集中存储的状态转移。根据我们最初的表述，还需要考虑这些分布外动作。

在奖励函数本身的定义中，分布外的不真实行为被标记为无效；如果智能体在学习过程中采样到该行为，

[2]受篇幅限制，必要的背景知识和算法概述请参阅 [47] 。

就会导致当前回合终止。换言之，错误的猜测（奖励为 0）会导致进入终止状态。这一简单约束迫使估计误差最小化，从而避免产生潜在的估计偏差。因此，导致当前回合终止的不真实行为，可以避免时序差分目标中错误的自举回报估计在未来不断传播。此举旨在尽量减少智能体策略与行为策略之间差异所导致的分布偏移问题。更具体地说，它明确确保无论下一个采样到的动作是什么，当前策略 $\pi(a|s_t)$ 都尽可能接近行为分布 $\pi_\beta(a|s_t)$。尽可能少地在分布外且未见过的动作上查询 Q函数，因为这最终会增加估计中的错误。

这种错误，即“外推误差” [22]，是对状态-动作对进行不切实际且错误的估计时产生的。当选取估计值 $Q(s,\ a)$中的动作 a_t时，后续的状态-动作对（$s_t a_t$）因无法在数据集中找到而与数据集不一致，从而导致这种错误。它会带来噪声，进而造成持续的高估偏差；在离策略设置下，由于无法收集新数据，这种偏差无法得到纠正 [21, 22]。在离线设置下直接使用深度Q网络，可能导致性能下降，并出现类似过拟合的情况 [38]。我们提出的机制能够最大限度地减少这些错误。需要注意的是，与基于行为克隆的方法不同，我们不会强行将“正确”动作输入智能体。我们让智能体像在与环境实时交互时一样自主做出确定性决策，从而保留原始算法的整体工作流程。这提供了一个统一的接口，可轻松地在离线学习与在线学习之间切换。

最后，需要指出的是，本研究旨在加深我们对人们为何跳过音乐的理解，并找出用于检测该行为的高质量特征。为此，我们分析了深度强化学习在预测这一行为方面的适用性。我们将进一步针对音乐跳过预测任务调整该方法，并使用近期提出的离线无模型算法进行评估[4, 13, 20, 33] ，留待未来工作完成。尽管如此，我们提出的方法无需进行任何架构或算法上的修改。它有望实现在线学习与离线学习之间的快速切换，反之亦然。此外，它也可被视为对智能体进行快速预训练，之后可将其部署到线上进行持续学习。

4 实验设置

4.1 数据集

我们使用 Spotify [11] 提供的真实世界音乐流媒体会话数据集（MSSD）开展实验。公开的训练集包含约 1.5 亿条流媒体会话日志，这些日志是在 2018 年 7 月 15 日至 9 月 18 日期间收集的，历时 66 天。每天包含十份日志，每份日志都纳入了全天范围内随机均匀抽样的音乐流媒体会话。每个会话包含 10 至 20 条记录，定义为用户听过的歌曲/曲目序列（每首歌曲对应一条记录）。每条记录都包含有关流的各种用户上下文信息（例如播放列表类型），以及用户与平台的交互历史（例如拖动进度，即在曲目内向前/向后搜索的次数）。尽管没有提供曲目标题，但提供了描述性的音频特征和元数据（例如原声度、效价和发行年份）。需要注意的是，数据中没有用户身份标识，也无法获取人口统计或地理信息。因此，由于无法判断两个会话是由同一用户还是不同用户播放的，本研究着重于对用户跳过行为的建模与理解。

4.1.1 时间相关性。 聆听会话之间不存在时间相关性，即这些会话并非按历史顺序呈现，这一点体现在相邻会话可能发生在一天中差异很大的时段（例如早晨和晚上）。此外，给定日期内的十条日志也没有顺序之分（即，第一天的第 1 条日志不一定早于同一天的第 2 条日志）。这并不妨碍深度强化学习应用于跳过预测任务的可能性，因为系统提供了歌曲播放时段。因此，可以对依赖于一天中不同时段的跳过行为进行建模。

4.1.2 训练集和测试集的创建。 在这项工作中，我们仅使用训练集，因为测试集中未提供评估时用作真实值的大多数元数据和跳过属性。通过从原始训练集中选取日志，我们训练集和测试集的统计信息如表 1 所示。从统计数据可以看出，所有集合中跳过值为 True 和 False 的比例均衡。这种均衡分布是该数据集及所有可用日志的固有属性。由于数据量庞大，因此对计算资源和执行时间的要求较高，我们使用第一个可用日期的前四条日志进行训练。我们在多条日志上进行测试，以检验模型对不同日期数据的泛化能力。除 T1 外，其他日志的索引、日期和／或月份均为随机选取；T1 是训练集采集后紧接着的第 5 条连续日志。采用随机选择方法的依据是，同一天的日志之间不存在时间相关性。这样做是为了展示我们提出的方法的泛化能力，并全面分析用户历史数据的重要性。

4.1.3 数据预处理。 除跳过特征、会话和歌曲标识符外，状态表示包含所有可用特征，其完整说明见 [11], 。类别特征（例如播放列表类型，以及导致当前曲目开始播放或结束播放的用户动作）采用独热编码。所有音频特征均经过标准化，使其分布的均值为 0，标准差为 1。总的来说，状态表示由 70 个特征组成。为便于讨论，这些特征分组如下：

用户行为 (UB)：

- **结束原因（RE）**是导致当前播放结束的原因。这是一个独热编码特征，因此涵盖了多种编码特征，例如 *Trackdone*、返回按钮、*Fwdbtn*和结束播放。- **开始原因（RS）**。与结束原因类似，它指的是导致当前播放开始的动作类型。

表 1：预处理后用于实验的数据集汇总。log(s) # 表示从可用的十个日志中选取的日志。skip (%) 指 True 和 False 值的比例。

数据集	Date	日志数 #	记录数 #	会话数	跳过（%）
训练集	2018年7月15日	[0, 3]	11,927,861	711,838	51.20%
测试集 (T1)	2018年7月15日	4	2,991,438	178,419	51.21%
测试集 (T2)	19/07/2018	8	3,395,883	204,145	50.53%
测试集 (T3)	27/07/2018	0	3,447,209	207,060	50.76%
测试集 (T4)	10/08/2018	6	3,407,685	205,267	50.42%
测试集 (T5)	09/09/2018	1	2,588,711	155,617	51.48%

• **暂停（PA）**是两次播放之间暂停的时长。它分为无、短和长暂停。
• **拖动进度 (SC)** 是播放过程中向前或向后跳转的次数。它们分别对应快进次数 和回退次数。

• **播放列表切换 (PS)** 表示用户是否在当前播放过程中更改了播放列表。

上下文 (CX):

• **会话时长 (SL)**是聆听会话的时长。
• **会话位置（SP）**是曲目在会话中的位置。

• **一天中的时段（HD）**是发生播放时的小时（[0..23]）。

• **播放列表类型（PT）**是发生播放时所在播放列表的类型。例如：用户收藏、个性化播放列表和广播。

• **高级版（PR）**表示用户是否使用高级版。

• **随机播放（SH）**表示播放曲目时是否启用了随机播放模式。

内容（CN）。第三类也是最后一类，汇集了所有**曲目（TR）**的元数据和特征，因为它们是 MSSD 中唯一基于内容的信息。它包含 28 个特征，例如节拍强度、调性、时长以及八个声学向量（[0..7]）。

4.2 评估指标

为评估我们提出的方法，我们采用了*Spotify* 序列跳过预测挑战赛中的评估指标。这也有助于与选定的基线方法进行公平比较，因为这些方法正是针对该挑战赛及以下任务提出的：给定一个聆听会话，预测某位用户是否会跳过会话后半段中遇到的各个曲目。因此，我们使用选定测试集中的每个会话后半段进行预测。如果一个会话包含奇数条记录，则将中间值向上取整。这样做是因为，准确反映用户紧接此前的交互，有助于音乐流媒体服务生成未来的推荐。因此，推断用户是否会跳过当前曲目以及会话中的后续曲目都很重要。我们采用首次预测准确率和平均准确率作为评估指标。

首次预测准确率（FPA）是对每个会话后半段中的首次交互进行预测时的准确率。

平均准确率均值（MAA）定义如下：

$$MAA = \frac{\sum_{i=1}^{T} A(i)L(i)}{T} \tag{1}$$

其中，T 表示给定会话中待预测的曲目数量，$A(i)$表示序列截至位置 i 的准确率，$L(i)$表示 i^{th} 预测是否正确。直观而言，在这些评估指标中，越早做出的预测所获得的重要性越高。然而，在我们的设置中，问题表述并未利用这一特性。相反，我们指示智能体优化会话中正确预测的总数量。这样做是为了使系统的规范保持简单，并且能够轻松适配不同的指标和/或任务。
在数据集模式中，预测基于跳过_2特征。该特征表示一个阈值，用于判断用户是否只短暂播放了该曲目（未提供确切阈值），随后便跳到其会话中的下一首歌曲。

4.3 模型

4.3.1 基线方法。 为了确定音乐跳过预测任务中最先进的基线方法，我们对以 MSSD 数据集为基础的既有研究进行了广泛检索。我们确定了 Spotify Sequential Skip Pre- diction Challenge 中排名前五的提交方案里排名前四的方案，并在 WSDM Cup 2019 研讨会上进行了介绍：

• **多任务 RNN**：基于 RNN 的方法，用于预测多种隐式反馈（多任务） [68]。
• **Multi-RNN**：由两个不同的堆叠 RNN 组成的 Multi-RNN，其中第二个 RNN 根据第一个 RNN 做出跳过预测；第一个 RNN 充当编码器 [23]。
•**时序元学习**：一种序列学习和元学习方法，由膨胀卷积层和高速公路激活函数组成 [12]。

• **加权 RNN**：一种采用双层堆叠 LSTM 层，并通过加权损失函数进行训练的 RNN 架构 [31]。

它们在 Spotify 挑战赛中分别取得了第 1、第 2、第 3 和第 5 名的最佳整体表现，其中多任务 RNN 是最强的基线方法，加权 RNN 则是最弱的基线方法。我们在评估中未纳入该挑战赛整体排名第 4 的最佳模型，原因是未找到相关论文和代码仓库。对于选定的基线方法，我们使用随论文提供的代码（引用论文中提供了 GitHub 链接）。随后我们

尽我们所能，并采用经过优化的参数集，通过运行其提供的公开代码，在本地复现了它们的结果。然而，尽管我们尽了最大努力，所得结果仍持续差于 Spotify 挑战赛公开排行榜和／或相关论文中报告的结果。挑战赛使用的测试集并未完全公开，也没有真实标注，因此无法进行本地评估。不过，鉴于我们创建训练集和测试集的流程（第 4.1.2 节），即在首个可用日期进行训练，并在不同日期／月份进行评估，我们有充分理由假设，我们选定的测试集与公开挑战赛所用测试集的整体数据分布相似。为公平比较，我们因此报告公开排行榜上的结果，因为这些结果优于本地评估所得的结果。

4.3.2 DQN 架构。 在本研究中，我们探索了九种最先进的 DQN 架构。我们遵循所提出的框架，对这些架构在用户音乐跳过行为预测任务中的表现进行了全面研究。它们包括基础型 [47], 双重型 [59], 对抗型 [63], 架构，以及各自的 n 步学习变体 [46]。我们还探索了部分可观测架构，包括观测堆叠 [47] 、基于门控循环单元（GRU）和长短期记忆（LSTM）的架构 [26]。

受篇幅限制，本文未报告这些架构变体之间的比较结果。不过，我们注意到，原始深度 Q 网络的表现最佳。这是因为它的性能相当，同时其架构显著简单，复杂度更低。因此，本文仅报告基础型 DQN 架构（下文简称“DQN”）的结果。

4.4 实验流程

我们使用以下参数训练深度Q网络：经验回放记忆容量为 10000，批次大小和更新频率均设为 256，学习率为 0.001，折扣因子为 0.9。策略网络由三个大小为 128 的全连接层和一个大小为 2 的动作价值线性输出层组成。该输出层用于计算每个动作的 Q 值。我们通过随机搜索和树结构帕岑估计器搜索选择超参数，并选出表现最佳的一组参数，在测试集合上进行评估。深度Q网络智能体通过 Tensorforce [32] 库实现。为确保研究结果完全可复现，相关代码已发布于
https://github.com/NeuraSearch/Spotify-XRL-Skipping-Prediction

为探究训练期间可能出现的不稳定性和发散现象，我们在每个测试集上运行所提出的深度Q网络方法五次。报告的结果为所有测试集上的平均值。最后，在训练阶段，由于学习受到分布外动作的限制，数据集中的某些状态价值对不会被智能体经历，这是因为回合提前终止。在测试阶段，系统按顺序检索所有回合记录，智能体则在完整回合上确定性地执行动作，以进行评估。

4.4.1 事后分析。 为了分析用户历史数据在预测音乐跳过行为方面的重要性和有效性，我们首先采用 Shapley 加性解释框架（SHAP） [41]。这是一种基于博弈论的用于解释机器学习模型预测结果的方法。具体而言，我们采用 Kernel 解释器，这是一种与模型无关的 SHAP 值估算方法。这是因为目前没有专门针对深度强化学习的解释器。然而，由于 Kernel 解释器对待解释预测的模型不作任何假设，因此计算开销非常大。这意味着它比其他针对特定模型类型的算法更慢。考虑到这些巨大的计算需求，对于每个测试集，我们估算前 50 个回合（即聆听会话）的特征重要性值，每条记录使用 200 个扰动样本。鉴于所有测试集之间的相似度很高，我们仅报告 T1 的结果。

4.4.2 消融分析。为验证 SHAP 结果，我们对输入状态表示进行了消融分析。我们研究类别（例如，用户行为）和类型（例如，推荐系统）特征对深度Q网络性能的影响。为此，我们在不包含所选特征类型的状态表示上训练并评估所提出的方法（遵循上述相同的实验流程）。这种迭代方法每次评估仅移除一种类型，并重复进行，直至评估完构成输入状态表示的所有类型。

4.4.3 时间数据泄漏。对 MSSD 数据集进行更深入的调查，并结合事后分析和消融分析的验证，发现某些特征存在时间数据泄漏。这些特征此前未被注意到，并无意间影响了 Spotify 挑战赛，进而影响了基线方法。这些特征对应于会话长度（*SL*）以及导致当前播放结束的用户操作（*RE*）。这是因为它们向模型提供了来自未来的信息，而这些信息在实时预测系统中本不应可用。尽管我们认识到并承认这是一个问题，但与所选基线方法进行比较时报告的结果并未移除这些特征。这样做是为了与所选基线方法进行公平比较，因为它们在输入表示中包含这些特征。在研究人们为何跳过音乐时，我们会从状态中移除这些特征，并将其称为“修正后”状态。

5 结果

首先，我们通过与最先进的深度学习模型进行比较，证明了我们预测用户音乐跳过行为的方法是有效的。我们对音乐跳过预测任务和 MSSD 数据集的分析揭示了一个时间数据泄漏问题（第 4.4.3 节）。通过移除这些特征来“修正”状态表示后，我们全面考察了如何通过分析 UB、CX 和 CN 的重要性来检测跳过行为。

5.1 深度强化学习对音乐跳过预测的适用性（研究问题1）

在我们的本地评估中，尽管 Multi-RNN 和时序元学习在挑战赛提交结果中优于加权 RNN，但它们在我们选取的测试集上的表现始终更差。多任务 RNN 是公开挑战赛中表现最佳的基线方法，其表现

[3]排行榜结果见引用的论文和/或 https://www.aicrowd.com/challenges/spotify-sequential-skip-prediction-challenge

表 2：我们提出的深度Q网络方法与基线方法的 MAA 和 FPA 结果。所报告的结果是深度Q网络在所有测试集上的平均值（含 95% 置信区间）。对于基线方法，我们报告 Spotify 挑战赛[3]中公开可用的结果。这是为了进行公平比较，因为这些结果优于我们本地评估所得的结果。由于无法获得基线方法的置信区间，因此未予报告。表现最佳的模型以粗体标出。

	MAA		*FPA*	
	Mean	95% 置信区间	Mean	95% 置信区间
DQN	**0.820**	[0.818 - 0.822]	**0.881**	[0.880 - 0.882]
多任务 RNN	0.651	—	0.812	—
Multi-RNN	0.641	—	0.807	—
时序元学习	0.637	—	0.804	—
加权 RNN	0.613	—	0.794	—

略逊于加权 RNN。总体而言，我们注意到，所有基线方法在本地评估中的表现都始终不及公开挑战赛中的表现。对于多任务 RNN、Multi-RNN、时序元学习和加权 RNN，MAA 和 FPA 指标的表现分别下降了 4.9、16.2、4.9、0.8 (%) 和 2.4、8.2、2.2、0.4 (%)。因此，在表 2 中，我们报告了所提出的深度Q网络方法在 MAA 和 FPA 指标上的结果，并列出 Spotify 挑战赛公布的基线结果。这是因为这些结果优于我们在本地评估中获得的结果，也能提供尽可能公平的比较。我们提出的方法在 MAA 和 FPA 两项指标上均显著优于所有基线方法。与公开挑战赛中表现最佳的基线方法多任务 RNN 相比，我们提出的深度Q网络在 MAA 和 FPA 指标上的表现分别提升了 17% 和 7%。

总的来说，我们的结果证明了深度强化学习用于预测用户音乐跳过行为的有效性和适用性。基础型 DQN 架构能够胜过更复杂的、基于深度学习的最先进模型。此外，结果以及因篇幅限制而未在本文中呈现的深入分析也表明，只需显著更少的回合即可实现收敛，约为 2×10^5（训练集回合数的 ~ 1/4）。这表明我们提出的方法具有较高的样本效率，且能够快速收敛。因此，该方法也解决了深度强化学习广为人知的问题，即计算开销大、学习速度慢。与所选基线方法相比，我们的方法收敛迅速，且无需使用图形处理器。多次运行以及学习过程中性能的低波动性，也表明该方法学习稳定且有效。

5.2 时间数据泄漏的识别

在上一节中，我们将所提出的深度Q网络与选定的基线方法进行了比较，以验证该模型的有效性。通过尽可能公平的比较，实证结果表明我们的方法更具优势。然而，这种基准测试给模型引入了错误。这是因为，正如第 4.4.3 节所述，我们认识到 MSSD 中存在数据泄漏特征。*SL* 会告知模型某个用户将聆听多少首歌曲。但不应向模型提供这一信息，因为我们无法预知用户在当前聆听会话中会聆听多少首歌曲。此外，*RE* 特征提供了当前流如何结束的信息。也不应向模型透露这一信息。然而，为了与基线方法进行公平比较，尽管我们意识到了这一点，但由于这些特征包含在它们的输入表示中，因此并未将其删除。

图 1 验证了时间数据泄漏问题。该图报告了输入状态表示中所有特征对模型输出（SHAP）的平均影响分析。可以看到，用于检测音乐跳过行为的区分度最高的特征是 *RE Trackdone*，其次是 *RS Trackdone*、*RS Fwdbtn* 和 *Short PA*。*SL* 也被发现具有一定影响（排名第 19）。显然，所提出的深度Q网络认为这些特征质量很高，对于预测用户的音乐跳过行为至关重要。然而，这些特征会导致数据泄漏问题。将它们从输入状态表示中移除后，我们发现所提出的深度Q网络的性能有所下降：MAA 和 FPA 分别降低了 16% 和 11%。此外，对于多任务 RNN、Multi-RNN、时序元学习和加权 RNN，MAA 的性能分别下降了 5.2%、26.2%、7.6% 和 0.6%，FPA 的性能分别下降了 3.5%、28.4%、6.0% 和 1.3%（差异是将本地评估中移除这些特征后得到的结果与公开挑战赛报告的结果进行比较后计算得出的）。总体而言，这些结果验证了我们的初步直觉，并揭示了数据泄漏问题。这一发现强烈表明，未来应建立审慎的数据收集流程，以透明地衡量用户行为。离线基准应尽可能真实地反映现实世界中的（在线）任务。

5.3 用户行为、上下文和内容在检测跳过音乐中的作用（研究问题 2）

在本节最后部分，我们旨在回答主要研究问题：人们为什么跳过音乐？为此，我们识别并移除了状态表示中的数据泄漏特征，以便正确建模用户的音乐跳过行为。

5.3.1 用户行为（***UB***）。图 2 展示了所提出的深度Q网络在“修正后”状态表示上的 SHAP特征重要性分析。可以观察到，用户通过何种方式与底层平台交互以开始当前播放（即*RS*类型），被认为是检测跳过音乐行为时区分度最高的特征。*Trackdone* 和*Fwdbtn* 分别是预测跳过行为时负相关和正相关程度最高的特征。

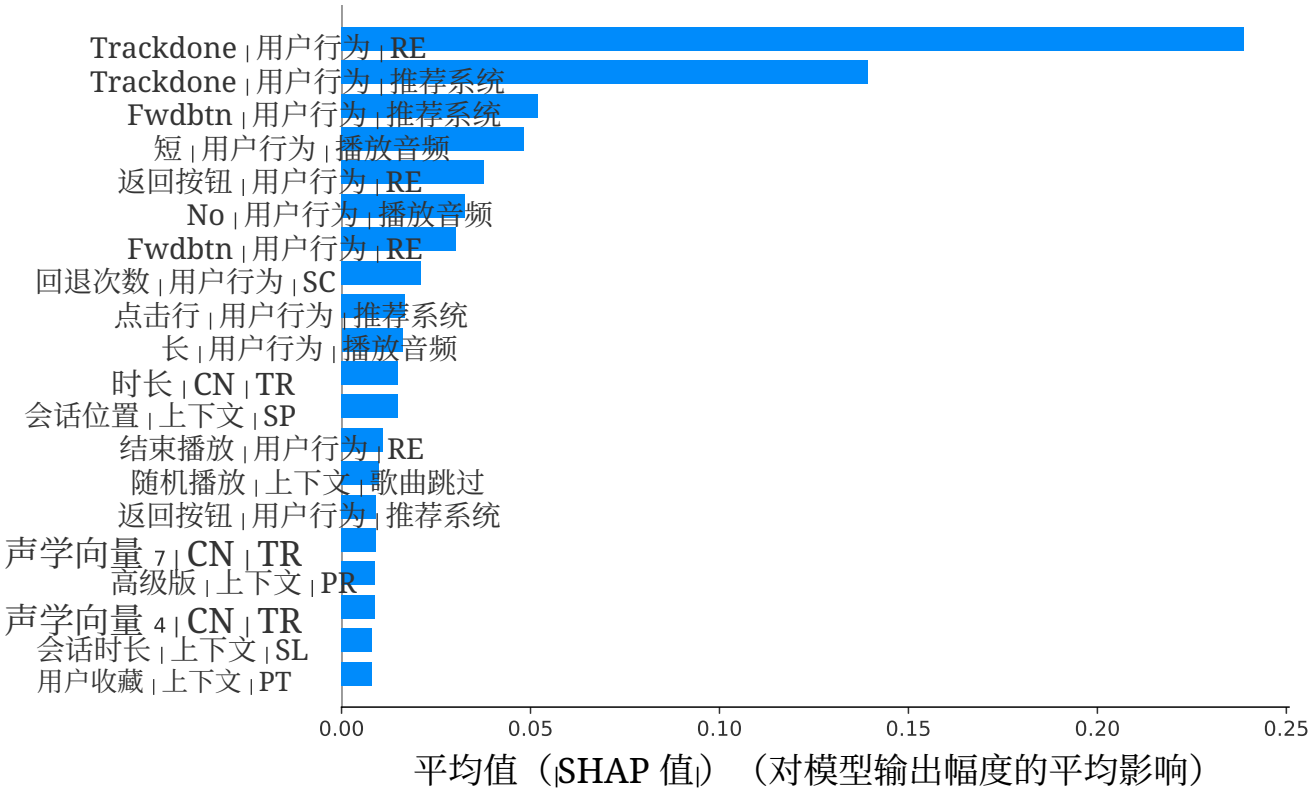

图 1：所提出的深度Q网络的 SHAP 特征重要性分析。特征的分类以及所用首字母缩略词的解释见第 4.1.3 节。特征按重要性排序，并以"[名称] [类别] | [类型]"的格式呈现。

它们分别对应于用户完整听完上一段播放内容，或在上一段播放过程中按下前进按钮（即跳过），然后开始当前播放。这些发现验证了 Meggetto 等人 [44] 最近的观察结果。基于他们定义的听众和跳过者用户类型，我们假设，能够反映用户属于这两种类型之一的用户行为是*RS Trackdone* 或*Fwdbtn*。从我们的结果可以清楚看出，用户与上一首歌曲的交互方式似乎极大地影响了深度强化学习检测其下一步交互方式的能力。另一种影响显著的用户行为是两次播放之间的暂停。研究显示，*Short PA*和*No PA* 分别强烈和轻微地暗示用户会跳过音乐。对于 *Long PA*，我们的结果强烈表明用户不会跳过当前歌曲。这一发现验证了我们的初始假设。这可能对应于用户在曲库中寻找想听的歌曲，因此暂停时间较长。因此，用户不太可能跳过该歌曲是合乎情理的。然而，短暂停顿对检测跳过音乐行为的影响令人意外。这或许可以用用户处于探索状态来解释：他们浏览曲库，并短暂试听多首歌曲，直到找到符合自身需求的歌曲。

5.3.2 上下文 ***(CX)***。 我们观察到，以随机播放 模式收听和/或使用高级版 账户的用户，跳过活动较少。使用用户收藏 播放列表收听则与较高的跳过率相关。研究还表明，在个性化播放列表 或广播中收听时，收听时间更长，因此跳过活动较少。这一发现可能表明这些用户的参与度更高。然而，这一点无法量化，仍需进一步评估以理解这一现象。跳过活动本身具有噪声性，且音乐可能被被动地（作为背景音乐）收听，例如收听广播时，这些因素或许可以解释这一现象。尽管播放列表 的研究结果似乎部分验证了以往工作 [44]，但在消融分析中，我们发现移除这些因素后，状态表示并未对深度强化学习预测音乐跳过的能力产生显著影响。

5.3.3 内容（***CN***）。MSSD 中唯一基于内容的特征与用户正在聆听的曲目（*TR*）相关。跳过活动与*TR* 特征之间的相关性不太明显，因为这些特征在检测音乐跳过时的区分度和显著性似乎较低。节拍强度和调性虽然大多集中在零影响附近，但表明较强的节拍强度与更多聆听相关，而高音调歌曲（调性）则更容易被跳过。此外，较长的歌曲（时长）通常与较高的聆听活动相关，但也可能对应跳过行为。然而，在消融分析中，我们观察到移除所有*TR* 特征不会影响深度Q网络的性能。我们认为这一结果出乎意料，因为这似乎

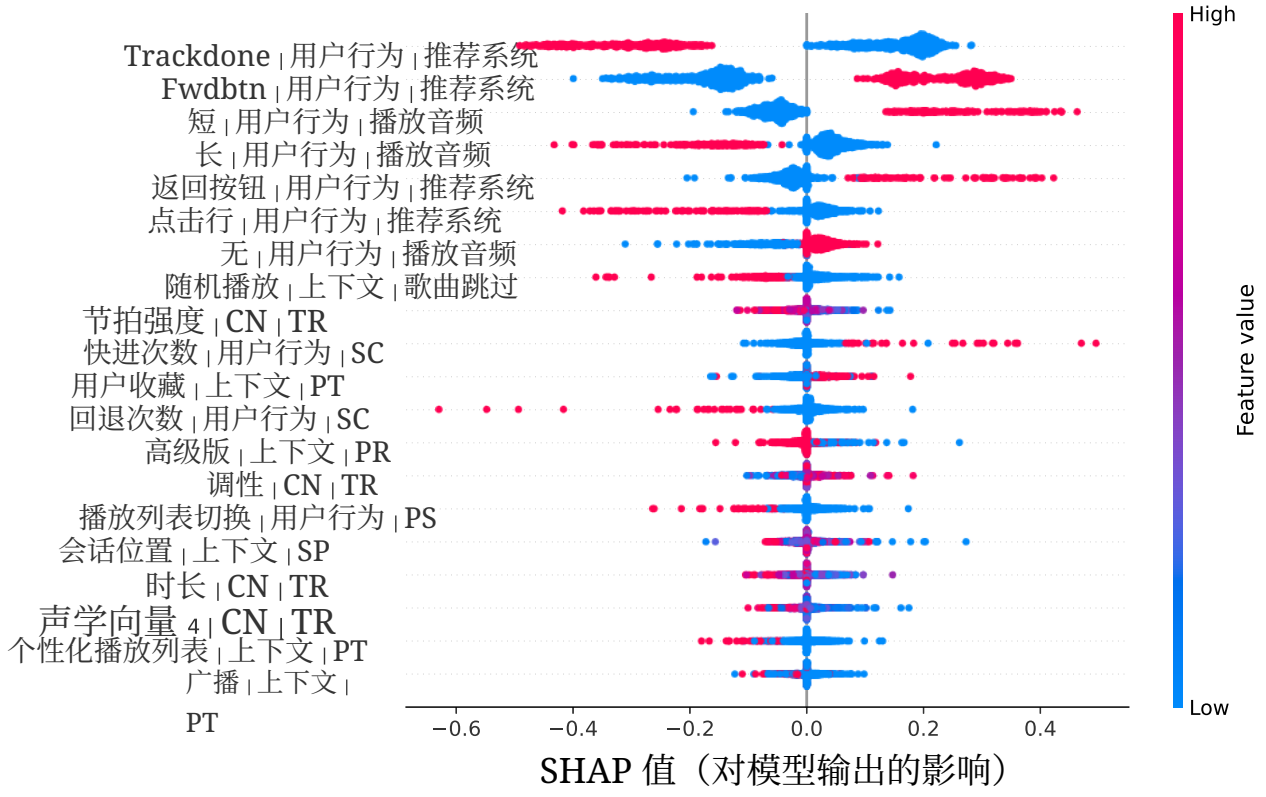

图 2：针对经过“修正”的状态表示（即解决时间数据泄漏后），所提出的深度Q网络在具有正向（跳过）和负向（未跳过）影响值方面的 SHAP 特征重要性分析。特征值轴表示观测值高低。对于布尔特征（例如，*RS Trackdone*），高/红色表示 True 值，低/蓝色表示 False 值。特征分类及所用首字母缩略词的说明见第 4.1.3 节。特征按重要性排序，并以“[名称] | [类别] | [类型]”的形式列出。

表3：在校正后的状态表示上，对所提出的深度Q网络进行消融分析所得的 MAA 和 FPA 结果。所报告的结果为所有测试集的平均值及 95%置信区间。（*）和（）表示所选特征类型对所提出的深度Q网络（基于“校正状态”）在 MAA 或 FPA 上的性能具有统计显著影响。这分别基于置信水平（$p < .05$）和（$p < .001$）。**

	MAA		*FPA*	
	Mean	95% 置信区间	Mean	95% 置信区间
校正状态	0.664	[0.662 - 0.666]	0.773	[0.772 - 0.774]
推理起始（RS）	0.389（**）	[0.378 - 0.400]	0.479（**）	[0.464 - 0.494]
暂停（播放音频）	0.659（*）	[0.657 - 0.661]	0.769（*）	[0.768 - 0.770]
拖动进度（SC）	0.659	[0.655 - 0.663]	0.770（*）	[0.768 - 0.772]
播放列表切换（PS）	0.662	[0.659 - 0.665]	0.773	[0.772 - 0.774]
一天中的时段（HD）	0.663	[0.661 - 0.665]	0.773	[0.772 - 0.774]
播放列表类型（PT）	0.663	[0.661 - 0.665]	0.772	[0.771 - 0.773]
高级版（PR）	0.664	[0.662 - 0.666]	0.773	[0.772 - 0.774]
随机播放（SH）	0.663	[0.660 - 0.666]	0.774	[0.773 - 0.775]
曲目 (TR)	0.664	[0.661 - 0.667]	0.773	[0.772 - 0.774]

这与先前认为音频特征会影响人们跳过音乐方式的研究相矛盾 [14, 49]。

5.3.4 消融分析。 为验证我们的发现，并展示这些特征对深度Q网络性能的影响（无论该影响在统计上是否显著），我们在表3中报告了消融分析的结果。我们针对所提出的深度Q网络在每种选定类型的特征（例如，推荐系统）下的预测准确率进行了配对t检验，其中输入状态表示经过了“修正”。我们使用（*）和（**）表示，移除所选特征类型会对所提出的深度Q网络在MAA和FPA上的表现产生统计显著影响。这分别基于置信水平（$p < .05$）和（$p < .001$）。如先前图 2 所示，推荐系统 特征类型是检测跳过音乐的最佳质量估计指标。移除该特征后，MAA和FPA的性能分别下降28%和29%。播放音频也表现出显著影响。其余所有特征（包括上下文和CN类别）似乎都不会对深度Q网络的性能产生统计显著影响。因此，这些结果表明，确实可以利用数量有限的用户数据来预测用户的音乐跳过行为，且只有推荐系统 和播放音频 用户行为表现出统计显著影响。

6 讨论与结论

在这项工作中，我们旨在理解人们为何会跳过音乐。为开展此类分析，我们首先提出将深度强化学习用于歌曲聆听会话中用户跳过行为的序列预测任务。通过先了解深度强化学习模型如何学习个体用户行为，我们便能够帮助解释基于深度强化学习的音乐推荐系统所作出的推荐。为此，我们拓展了深度强化学习在这一分类任务中的适用性。在真实世界的音乐流媒体数据集（Spotify）上的结果表明，我们的方法有效：在 MAA 和 FPA 指标方面优于最先进的深度学习模型（**研究问题1**）。通过实证展示所提方法的有效性，我们的主要事后分析和消融分析围绕一项全面研究展开，探讨用户历史数据在所提深度强化学习模型检测音乐跳过行为时的效用和影响（回应**研究问题 2**）。

我们的研究结果表明，用户与平台的交互方式是准确检测跳过行为（即推荐系统 和播放音频）最具区分力的指标。令人意外的是，本研究探讨的聆听上下文和 CN 特征似乎不会影响用于预测音乐跳过行为的深度强化学习模型。我们的分析还揭示了数据集中部分特征所导致的时间数据泄漏问题；这些特征也用于公开挑战赛，因其包含未来信息，不应提供给实时预测系统。总而言之，本研究表明，只需利用有限的用户数据，便能准确表征用户的跳过行为。这对设计新颖、以用户为中心的音乐推荐系统具有重要启示：通过精简并筛选高质量数据特征，避免引入错误和偏差。研究结果及对所提方法的深入分析表明，该方法具有较高的样本效率、快速收敛和长期稳定性。由于只需少量回合即可实现收敛，因此可通过提前终止大幅缩短训练时间。该方法也不需要使用图形处理器（与基于最先进深度学习的模型不同），我们的方法还明确解决了深度强化学习计算开销巨大的这一众所周知的局限。这些发现以及始终稳定的性能，使本研究对未来研究具有重要价值。

鉴于对用户跳过行为进行建模和理解的重要性，我们认为这项工作是改进用户建模技术的重要一步。对跳过行为进行准确表征，能够为底层推荐过程提供极有价值的信息流。例如，我们预计，我们的发现（如推荐系统类型）与下游任务密切相关，即实时捕捉用户的跳过类型 [44]。通过拓展我们的方法来预测并理解其他用户行为，我们可以全面表征听众的偏好、兴趣和需求。我们还主张，在收集数据并将其提供给模型以衡量用户行为时，应审慎考量。随着人们日益关注用户数据收集与隐私，尽可能减少数据收集至关重要。未来的工作可以拓展我们的方法，以预测何时歌曲可能会被跳过。这类信息有助于预测歌曲中最可能发生跳过行为的时刻，这对底层平台可能极具价值。进一步研究用户的情绪或当前心理状态如何影响其跳过行为，也很有意义。若能获取更丰富的行为数据和未经匿名化处理的聆听会话，另一条研究方向可以探讨跳过信号与个体用户偏好之间的关系（例如，情境感知多媒体推荐系统）。最后，尽管提升性能并非本研究的目标，但针对音乐跳过预测任务进一步调整我们的方法，有望带来性能提升。鉴于本研究以用户为中心的探索性质，我们将进一步使用新兴的无模型离线强化学习算法和架构进行实验与评估（例如，将我们的分析拓展至基于Transformer的深度强化学习模型 [30]），留待未来研究。

致谢

本研究得到了工程与物理科学研究理事会的资助 [，资助编号为 EP/R513349/1]。

参考文献

[1] Sainath Adapa。2019。使用循环神经网络对会话进行序列建模以实现跳过预测。*arXiv* 预印本 *arXiv:1904.10273*（2019）。[2] Darius Afchar 和 Romain Hennequin。2020。利用归因方法提高神经网络的可解释性：应用于隐式信号预测。载于 第十四届*ACM*推荐系统会议。220–229。[3] M Mehdi Afsar、Trafford Crump 和 Behrouz Far。2021。基于强化学习的推荐系统：综述。*ACM*计算调查（*CSUR*）（2021）。[4] Rishabh Agarwal、Dale Schuurmans 和 Mohammad Norouzi。2020。对离线强化学习采取乐观视角。载于 国际机器学习会议。PMLR，104–114。[5] Snehasish Banerjee 和 Anjan Pal。2021。在 YouTube 上跳过可跳过广告：如何、何时、为何以及为何不跳过？载于 *2021*年第*15*届泛在信息管理与通信国际会议（*IMCOM*）。IEEE，1–5。[6] Daniel Belanche、Carlos Flavián 和 Alfredo Pérez-Rueda。2017。用户对互动广告形式的适应：先前接触、习惯和时间紧迫感对广告跳过行为的影响。远程信息处理与信息学34，7（2017），961–972。[7] Daniel Belanche、Carlos Flavián 和 Alfredo Pérez-Rueda。2020。YouTube 上可跳过与不可跳过广告的品牌回忆：针对活跃受众重新调整信息与唤醒度。在线信息评论44，3（2020），545–562。

[8] Ferenc Béres、Domokos Miklós Kelen、András Benczúr 等。2019。使用深度学习和集成方法进行序列跳过预测。（2019）。[9] Alex Beutel、Paul Covington、Sagar Jain、Can Xu、Jia Li、Vince Gatto 和 Ed H Chi。2018。潜在交叉：在循环推荐系统中利用上下文。载于《第十一届 ACM 网络搜索与数据挖掘国际会议论文集》。46–54。[10] Klaas Bosteels、Elias Pampalk 和 Etienne E Kerre。2009。利用电台日志和模糊集理论评估并分析动态播放列表生成启发式方法。载于 ISMIR，第 9 卷。351–356。[11] Brian Brost、Rishabh Mehrotra 和 Tristan Jehan。2019。音乐流媒体会话数据集。载于《万维网大会论文集》。2594–2600。[12] Sungkyun Chang、Seungjin Lee 和 Kyogu Lee。2019。流媒体音乐内容的少样本序列跳过预测。arXiv 预印本 arXiv:1901.08203（2019）。[13] Will Dabney、Mark Rowland、Marc Bellemare 和 Rémi Munos。2018。使用分位数回归的分布式强化学习。载于《AAAI人工智能会议论文集》，第 32 卷。[14] Jonathan Donier。2020。音乐流媒体平台上跳过行为的普遍性。arXiv 预印本 arXiv:2005.06987（2020）。[15] Gabriel Dulac-Arnold、Ludovic Denoyer、Philippe Preux 和 Patrick Gallinari。2011。逐数据分类：一种应对稀疏性的序列方法。载于《欧洲机器学习与数据库知识发现联合会议》。Springer，375–390。[16] Gabriel Dulac-Arnold、Nir Levine、Daniel J Mankowitz、Jerry Li、Cosmin Paduraru、Sven Gowal 和 Todd Hester。2021。现实世界强化学习的挑战：定义、基准测试与分析。《机器学习》（2021），1–50。[17] Jun Feng、Minlie Huang、Li Zhao、Yang Yang 和 Xiaoyan Zhu。2018。利用强化学习从噪声数据中进行关系分类。载于《AAAI人工智能会议论文集》，第 32 卷。[18] Andres Ferraro、Dmitry Bogdanov 和 Xavier Serra。2019。根据会话中曲目的声学特征，使用提升树进行跳过预测。arXiv 预印本 arXiv:1903.11833（2019）。[19] Benjamin Fields 等。2011。让聆听融入上下文：作为推荐引擎的播放列表。博士论文。戈德史密斯学院（伦敦大学）。[20] Scott Fujimoto、Edoardo Conti、Mohammad Ghavamzadeh 和 Joelle Pineau。2019。批量深度强化学习算法的基准测试。arXiv 预印本 arXiv:1910.01708（2019）。[21] Scott Fujimoto、Herke Hoof 和 David Meger。2018。解决演员-评论家方法中的函数近似误差。载于《国际机器学习会议》。PMLR，1587–1596。[22] Scott Fujimoto、David Meger 和 Doina Precup。2019。无需探索的离策略深度强化学习。载于《国际机器学习会议》。PMLR，2052–2062。[23] Christian Hansen、Casper Hansen、Stephen Alstrup、Jakob Grue Simonsen 和 Christina Lioma。2019。使用多重循环神经网络方法对序列音乐曲目跳过行为建模。arXiv 预印本 arXiv:1903.08408（2019）。[24] Casper Hansen、Christian Hansen、Lucas Maystre、Rishabh Mehrotra、Brian Brost、Federico Tomasi 和 Mounia Lalmas。2020。用于大规模音乐推荐的上下文与序列用户嵌入。载于《第十四届 ACM推荐系统会议》。53–62。[25] Christian Hansen、Rishabh Mehrotra、Casper Hansen、Brian Brost、Lucas Maystre 和 Mounia Lalmas。2021。将音乐流媒体平台上的消费转向多样化内容。载于《第十四届 ACM 网络搜索与数据挖掘国际会议论文集》。238–246。[26] Matthew Hausknecht 和 Peter Stone。2015。面向部分可观测马尔可夫决策过程的深度循环 Q 学习。载于《2015年AAAI秋季研讨会系列》。[27] Jaromír Janisch、Tomáš Pevnỳ 和 Viliam Lisỳ。2019。使用深度强化学习对高成本特征进行分类。载于《AAAI人工智能会议论文集》，第 33 卷。3959–3966。[28] Jaromír Janisch、Tomáš Pevnỳ 和 Viliam Lisỳ。2020。将高成本特征分类视为序贯决策问题。《机器学习》109，8（2020），1587–1615。[29] Dietmar Jannach、Massimo Quadrana 和 Paolo Cremonesi。2022。基于会话的推荐系统。载于《推荐系统手册》。Springer，301–334。[30] Michael Janner、Qiyang Li 和 Sergey Levine。2021。将离线强化学习视为一个大型序列建模问题。《神经信息处理系统进展》34（2021），1273–1286。[31] Olivier Jeunen 和 Bart Goethals。2019。使用基于会话的循环神经网络预测序列用户行为。（2019）。[32] Alexander Kuhnle、Michael Schaarschmidt 和 Kai Fricke。2017。Tensorforce：用于应用强化学习的 TensorFlow 库。网页。https: //github.com/tensorforce/tensorforce [33] Aviral Kumar、Justin Fu、Matthew Soh、George Tucker 和 Sergey Levine。2019。通过自举误差缩减稳定离策略 Q 学习。《神经信息处理系统进展》32（2019）。[34] Paul Lamere。2014。《跳过》。检索于2022年9月22日，网址：https:// musicmachinery.com/2014/05/02/the-skip/

[35] Paul Lamere。2015。《The Drop Machine》。2022年9月22日取自 https: //musicmachinery.com/2015/06/16/the-drop-machine/ [36] Sascha Lange、Thomas Gabel 和 Martin Riedmiller。2012。批量强化学习。载于《强化学习》。Springer，45–73。[37] Alexandra Langmeyer、Angelika Guglhör-Rudan 和 Christian Tarnai。2012。音乐偏好揭示了哪些人格特征？使用自我评定和音乐样本评分进行的跨文化重复研究。《个体差异杂志》33，2（2012），119。[38] Sergey Levine、Aviral Kumar、George Tucker 和 Justin Fu。2020。离线强化学习：教程、综述及开放问题展望。arXiv预印本 arXiv:2005.01643（2020）。[39] Enlu Lin、Qiong Chen 和 Xiaoming Qi。2020。用于不平衡分类的深度强化学习。《应用智能》（2020），1–15。[40] Feng Liu、Ruiming Tang、Xutao Li、Weinan Zhang、Yunming Ye、Haokun Chen、Huifeng Guo 和 Yuzhou Zhang。2018。基于深度强化学习并显式建模用户-物品交互的推荐。arXiv预印本 arXiv:1810.12027（2018）。[41] Scott M Lundberg 和 Su-In Lee。2017。解释模型预测的统一方法。神经信息处理系统进展30（2017）。[42] Coralie Martinez、Guillaume Perrin、Emmanuel Ramasso 和 Michèle Rombaut。2018。用于时间序列早期分类的深度强化学习方法。载于2018年第26届欧洲信号处理会议（EUSIPCO）。IEEE，2030–2034。[43] James McInerney、Brian Brost、Praveen Chandar、Rishabh Mehrotra 和 Benjamin Carterette。2020。通过序列奖励交互对列表推荐进行反事实评估。载于第26届 ACM SIGKDD知识发现与数据挖掘国际会议论文集。1779–1788。[44] Francesco Meggetto、Crawford Revie、John Levine 和 Yashar Moshfeghi。2021。音乐流媒体会话中的跳过行为类型。载于第30届 ACM信息与知识管理国际会议论文集。3333–3337。[45] Silvia Milano、Mariarosaria Taddeo 和 Luciano Floridi。2020。推荐系统及其伦理挑战。《人工智能与社会》35，4（2020），957–967。[46] Volodymyr Mnih、Adria Puigdomenech Badia、Mehdi Mirza、Alex Graves、Timothy Lillicrap、Tim Harley、David Silver 和 Koray Kavukcuoglu。2016。深度强化学习的异步方法。载于国际机器学习会议。PMLR，1928–1937。[47] Volodymyr Mnih、Koray Kavukcuoglu、David Silver、Andrei A Rusu、Joel Veness、Marc G Bellemare、Alex Graves、Martin Riedmiller、Andreas K Fidjeland、Georg Ostrovski 等。2015。通过深度强化学习实现人类水平控制。《自然》518，7540（2015），529–533。[48] Nicola Montecchio、Pierre Roy 和 François Pachet。2020。音乐流媒体服务用户的跳过行为及其与音乐结构的关系。《公共科学图书馆·综合》15，9（2020），e0239418。[49] Aaron Ng 和 Rishabh Mehrotra。2020。研究音频状态与状态转移对音乐流媒体会话中曲目排序的影响。载于第十四届 ACM推荐系统会议。697–702。[50] Elias Pampalk、Tim Pohle 和 Gerhard Widmer。2005。基于跳过行为的动态播放列表生成。载于 ISMIR，第5卷。634–637。[51] Peter J Rentfrow 和 Samuel D Gosling。2003。日常生活中的 do re mi：音乐偏好的结构及其人格相关性。《人格与社会心理学杂志》 84，6（2003），1236。[52] Peter J Rentfrow 和 Samuel D Gosling。2006。一首民谣中的讯息：音乐偏好在人际知觉中的作用。《心理科学》17，3（2006），236–242。[53] Markus Schedl、Peter Knees、Brian McFee 和 Dmitry Bogdanov。2022。音乐推荐系统：技术、应用场景与挑战。载于《推荐系统手册》。Springer，927–971。[54] Guy Shani、David Heckerman、Ronen I Brafman 和 Craig Boutilier。2005。基于马尔可夫决策过程的推荐系统。《机器学习研究杂志》6，9（2005）。[55] Yading Song、Simon Dixon 和 Marcus Pearce。2012。音乐推荐系统综述与未来展望。载于第9届计算机音乐建模与检索国际研讨会，第4卷。Citeseer，395–410。[56] Richard S Sutton 和 Andrew G Barto。2018。《强化学习：导论》。麻省理工学院出版社。[57] John R Taylor 和 Roger T Dean。2021。连续情感评分任务对陌生艺术音乐聆听时长的影响。《新音乐研究杂志》50，3（2021），242–258。[58] Charles Tremlett。2019。Spotify 序列跳过预测挑战赛的初步研究。（2019）。[59] Hado Van Hasselt、Arthur Guez 和 David Silver。2016。结合双重Q学习的深度强化学习。载于 AAAI人工智能会议论文集，第30卷。[60] Sergey Volokhin 和 Eugene Agichtein。2018。理解日常活动中的音乐聆听意图及其对上下文音乐推荐的启示。载于2018年人类信息交互与检索会议论文集。313–316。

[61] Shoujin Wang、Longbing Cao、Yan Wang、Quan Z Sheng、Mehmet A Orgun 和 Defu Lian。2021。基于会话的推荐系统综述。*ACM*计算调查（*CSUR*）54，7（2021），1–38。[62] Wenjie Wang、Fuli Feng、Xiangnan He、Liqiang Nie 和 Tat-Seng Chua。2021。为推荐任务去噪隐式反馈。载于第十四届 *ACM* 网络搜索与数据挖掘国际会议论文集。373–381。[63] Ziyu Wang、Tom Schaul、Matteo Hessel、Hado Hasselt、Marc Lanctot 和 Nando Freitas。2016。用于深度强化学习的对抗网络架构。载于国际机器学习会议。1995–2003。[64] Hongyi Wen、Longqi Yang 和 Deborah Estrin。2019。利用点击后反馈进行内容推荐。载于第十三届 *ACM* 推荐系统会议论文集。278–286。

[65] Marco A Wiering、Hado Van Hasselt、Auke-Dirk Pietersma 和 Lambert Schomaker。2011。用于解决分类问题的强化学习算法。载于*2011 IEEE* 自适应动态规划与强化学习研讨会（*ADPRL*）。IEEE，第 91–96 页。[66] Xiangyu Zhao、Long Xia、Liang Zhang、Zhuoye Ding、Dawei Yin 和 Jiliang Tang。2018。用于逐页推荐的深度强化学习。载于第 *12* 届 *ACM*推荐系统会议论文集。第 95–103 页。[67] Guanjie Zheng、Fuzheng Zhang、Zihan Zheng、Yang Xiang、Nicholas Jing Yuan、Xing Xie 和 Zhenhui Li。2018。DRN：用于新闻推荐的深度强化学习框架。载于*2018* 年万维网大会论文集。第 167–176 页。[68] Lin Zhu 和 Yihong Chen。2019。基于循环神经网络的基于会话的序列跳过预测。*arXiv* 预印本 *arXiv:1902.04743* (2019)。