**说 明 书**

**端上实时的视频流广告效果预估方法及广告位分配、广告计费方法**

**技术领域**

1. 本公开涉及视频流广告推荐，尤其涉及端上实时的视频流广告推荐效果的预估方法，以及最大化广告平台收益的广告分配、计费方法。

**背景技术**

1. 近年来，一种名为视频流广告的新型在线广告形式在一些移动电子商务平台和短视频应用程序上出现。在这一场景下，用户以信息流的方式浏览全屏广告，用户可以点击广告获取广告详情信息，或者向下滑动屏幕查看下一条广告。
2. 在全屏视频流场景中，用户将浏览各种产品，这导致用户的实时兴趣不断变化。因此，在视频流广告中一个基本问题是，广告推荐需要及时地响应用户的即时兴趣。
3. 端上重排技术通过在移动设备上部署重排序模型以捕获用户的实时兴趣。基于实时兴趣，端上重排序模型会重新评估剩余的候选广告并确定新的顺序，从而及时地响应用户的即时兴趣。然而，将这种技术应用于在线广告系统仍然存在两个关键挑战：一、广告间的相互影响会导致贪心广告分配策略不能达到最大化社会福利目标；二、缺乏经济学性质保证会导致广告市场的不稳定。
4. 由此期望提供一种基于端上重排技术的广告推荐效果预估方法，其能够更加准确地预估用户对广告的点击率、下翻率，显著提升用户体验。同时预估方法还预估广告所带来的期望社会福利，结合所提供的广告分配、计费方法解决上述关键挑战，达到最大化社会福利的目标，并且稳定广告市场。

论文引言：近年来，一种名为视频流广告的新型在线广告形式在一些移动电子商务平台和短视频应用程序上出现。在这一场景下，用户以信息流的方式浏览全屏广告，用户可以点击广告获取广告详情信息，或者向下滑动屏幕查看下一条广告。在全屏视频流场景中，用户将浏览各种产品，这导致用户的实时兴趣不断变化。因此，在视频流广告中一个基本问题是，广告推荐需要及时地响应用户的即时兴趣。

然而，由于传统在线广告平台在可扩展性和延迟方面的限制，展示的广告通常由部署在云服务器上的复杂信息检索系统决定。客户端然后通过分页请求机制从云端获取候选广告，之后按照云服务器预定的顺序显示它们。由于服务器传送的广告序列顺序是固定的，直接使用它们而不进行重排序可能无法满足用户的实时兴趣，导致广告性能下降。此外，由于网络延迟和数据隐私政策的原因，用户的实时行为特征无法被基于云的信息检索系统充分利用。一些近期的研究尝试在移动设备上部署重排序模型以捕获用户的实时行为。基于实时行为，设备上的重排序模型会重新评估剩余的候选广告并确定新的顺序。尽管设备上的重排序方法在实时推荐系统中显示出潜力，但将这种方法应用于在线广告系统仍然存在两个关键挑战。

第一个挑战来自全屏信息流场景中广告的外部性，这导致了当前广告显示机制的次优结果。以前的研究已经表明，物品之间的相互影响会导致用户对同一组候选广告集的不同排列有不同的偏好。在流媒体广告场景中，用户必须先查看前一则广告，然后才能滑动屏幕访问新的广告，这样的广告在用户心中留下了更深刻的印象。因此，这种相互影响的现象在这种场景中更为明显。因此，贪婪的逐点方法，即只显示有效千次展示成本（eCPM）最高的广告而不考虑其对候选广告的潜在影响，无法获得最大的预期社会福利。

第二个挑战在于广告拍卖机制中缺乏经济属性保证，这可能导致广告市场的不稳定。当我们设计一个考虑到广告外部性的新排序方案时，如果仍然坚持传统的拍卖而没有经过仔细设计，可能会导致无法保证经济属性。以广泛使用的广义第二价格拍卖（GSP）为例，GSP根据广告的eCPM以降序排列广告，并向每个广告主收取维持相同位置所需的最低出价。然而，由于设备上重排序的背景下，显示广告的eCPM并不是单调的，直接应用GSP可能会导致广告主的负效用。

在这项工作中，我们针对端上重排背景下的在线广告系统，提出了一种新颖的适应性广告拍卖（AdaAuc）机制，即适用于设备上的广告系统。具体来说，为了解决第一个挑战，我们将设备上的广告重排序建模为一个马尔可夫决策过程（MDP），这自然捕捉了广告外部性的特点：当前广告的决定可能会影响随后的广告。基于MDP的公式化，我们设计了一个新的分配规则，考虑到广告的外部性，以优化预期的社会福利。此外，我们界定了拍卖应满足的两个经济属性：事后激励兼容性（ex-post IC）和个体合理性（IR）。为了解决第二个挑战，我们设计了一种类似于VCG的支付规则，该规则使广告商和广告平台的目标一致。通过精心设计的分配和支付规则，AdaAuc能够在用户查看当前广告后适应性地选择合适的广告，并在用户点击时向广告商收费。我们证明了所提出的AdaAuc满足事后IC和IR，从而保证广告市场的稳定性。然后，我们采用基于强化学习（RL）的重排序模型来应用AdaAuc于广告系统。我们在一个公共数据集上进行了广泛的实验，比较了AdaAuc与文献中其他代表性基线的性能。这些实验结果表明，所提出的AdaAuc在预期社会福利方面能够优于其他基线方法。

**发明内容**

1. 本公开要解决的第一个技术问题是提供一种端上实时的视频流广告效果预估方案，其能够更加准确地预估用户点击广告和浏览下一条广告的概率，此外，为了广告分配、计费方法的实施，其还能准确预估展示某一广告带来的期望社会福利。
2. 根据本公开的第一方面，提供了一种信息推荐模型的训练方法，包括：
3. …
4. 根据本公开的第二方面，提供了一种广告的分配方法，包括：
5. …
6. 根据本公开的第三方面，提供了一种广告的计费方法，包括：
7. …
8. 根据本公开的第四方面，提供了信息推荐模型的训练装置，包括：
9. …
10. 根据本公开的第五方面，提供了广告的分配装置，包括：
11. …
12. 根据本公开的第六方面，提供了广告的计费装置，包括：

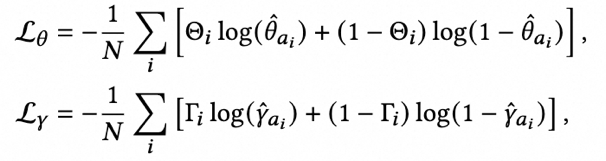
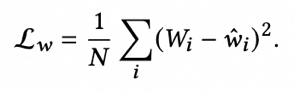
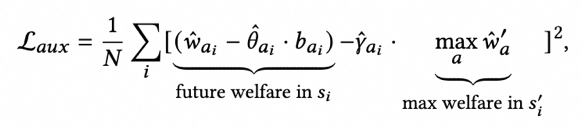
**技术效果**

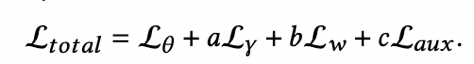
1. 本发明与现有技术相比，其优点表现为：使用了端上重排技术，其能够更加准确地预估用户对广告的点击率、下翻率，显著提升用户体验。设计了端上重排场景下的广告分配及付费方法，使广告平台能够最大化自身利益（最大化期望社会福利），也能保证IC、IR的经济学性质，从而避免广告主间策略性竞价行为，保障广告市场稳定性。

**附图说明**

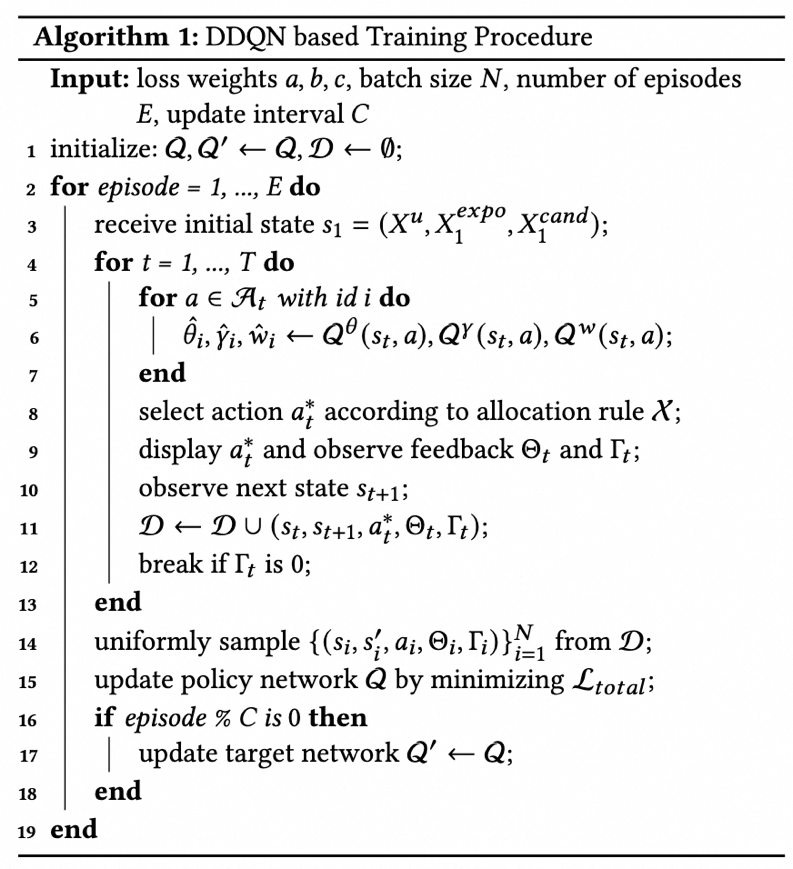
1. 图1是根据本公开一个实施例的基于强化学习的端上重排模型的示意性结构图。
2. 图2为不同广告分配方法的广告推荐效果的对比实验结果。
3. 图3为训练过程中不同广告分配方法的广告推荐效果的对比实验结果。
4. 图4为不同广告分配、计费方法下后验激励兼容性质的对比实验结果。

**具体实施方式**

1. 下面将参照附图更详细地描述本公开的示例性实施方式。虽然附图中显示了本公开的示例性实施方式，然而应该理解，可以以各种形式实现本公开而不应被这里阐述的实施方式所限制。相反，提供这些实施方式是为了使本公开更加透彻和完整，并且能够将本公开的范围完整地传达给本领域的技术人员。
2. **实施例1**
3. 下面，将说明本申请实施例在一个实际的应用场景中的示例性应用。
4. 在介绍具体模型及训练方法前，先明模型所需的用户和广告特征：
5. 用户资料是用户自行填写的个人信息，以及一些描述用户或端设备状态的信息，包括如下的至少一种特征：用户性别、年龄、地域等，及描述用户或端设备状态的额外特征，如电池电量、网络状况等。
6. 用户浏览记录，包括至少一条最近浏览的广告的浏览信息。浏览信息包括如下中的至少一种：广告侧特征，如广告的ID、广告投放者ID、广告的类别、广告所属的广告组、广告所属地域等，以及用户对广告的交互行为。
7. 候选广告集合，包括至少一条从云侧接收待展示的广告信息。广告信息包括如下中的至少一种：广告侧特征，如广告的ID、广告投放者ID、广告的类别、广告所属的广告组、广告所属地域、广告主竞价等。
8. 目标广告，需要预测其广告效果，包括如下的至少一种特征：广告侧特征，如广告的ID、广告投放者ID、广告的类别、广告所属的广告组、广告所属地域、广告主竞价等。
9. 这些特征在端设备上被实时维护，用于用户动态兴趣建模。实时维护这些特征对于捕捉用户即时的兴趣变化至关重要，因为它允许系统快速响应用户的最新兴趣。考虑到隐私问题，将特征数据保留在本地设备上而不是在云端处理，有助于保护用户的敏感信息，防止数据泄露或被不当使用。此外，本地处理数据还可以减少数据在网络中的传输时间，从而降低延迟，提供更加流畅和及时的用户体验。
10. 在上面介绍的特征说明基础上，下面首先结合具体的实施例对本公开提供的广告效果预估的模型装置进行介绍。
11. 首先结合图1进行介绍，图1为本公开实施例提供的广告效果预估的模型的实现示意图。
12. 如图1，构建的广告效果预估模型包括三个模块，特征提取器模块、基于MMoE多任务学习框架的用户偏好学习模块和基于多层感知机的社会福利学习模块，分别负责序列特征建模、用户反馈预测和社会福利预测。
13. 实施例中所述模型所需的用户和广告特征首先通过向量映射层，将离散或连续的特征映射成低维特征向量。而后这些特征向量被送入特征提取器模块。
14. 特征提取器模块：广告之间的相互影响会导致一个众所周知的现象，即同一组广告的不同排列可能会导致不同的广告投放效果。因此，设计了一个基于注意力的特征提取器，它包含两个多头注意力模块，用于提取用户浏览历史和候选广告集的上下文信息。为了建模用户浏览历史，我们从目标广告中移除了出价特征（图中的“过滤”框），因为用户的偏好不依赖于广告商的出价。具体地，我们对于用户浏览历史建模，我们将用户浏览历史的特征向量组作为键和值，将目标广告的特征向量作为查询，假设采用h头注意力模块以及浏览历史特征向量组中特征向量维度为，得到用户浏览历史向量表征为，其中待；同理，可以得到候选广告集的向量表征。
15. 这种方法可以帮助模型更好地理解用户在浏览广告时的行为和偏好，并相应地调整广告的排列，以改善广告效果，提供更个性化的用户体验。通过这种方式，系统能够利用用户的实时反馈和历史行为来优化广告投放，不仅提高了广告的相关性和吸引力，还能在保持用户满意度的同时最大化广告收入。
16. 用户偏好学习模块：用户反馈学习，包括学习用户对广告的点击率和下翻浏览下一条广告的概率，是一项监督学习任务。我们采用了广泛使用的多门控混合专家（MMoE）结构进行多任务学习。具体而言，它将用户资料特征向量、浏览历史向量表征以及目标广告特征向量作为输入，用于点击率和下翻率的预测。在这里，因为未展示的广告对用户兴趣不产生任何影响，候选广告集向量表征不作为模块的输入。
17. 社会福利学习模块：期望社会福利预测是一个强化学习任务。我们简单地使用多层感知机（MLP）来进行期望社会福利的预测。它将用户资料特征向量、浏览历史向量表征、候选广告集向量表征以及目标广告特征向量这些来自特征提取器的输入进行学习，以此来预测目标广告在给定状态下可能带来的最大预期社会福利。在社会福利学习模块中，我们纳入了候选广告集向量表征，以此来预估不同候选广告集下的期望社会福利。
18. 此外，本公开通过广告分配方法，达到最大化社会福利目的。通过设计付费方法，达到满足经济学性质，维护广告市场稳定的目的。
19. 广告分配方法：不同于已有广告分配方法一次性决定候选广告的展示顺寻，本公开提供的广告分配方法在云侧推荐系统返回了广告候选集合后，根据用户的实时兴趣，逐坑分配广告，从而最大化期望社会福利。具体来说，广告效果预估模型由于部署在端设备上，所以可以通过对用户的实时行为建模，从而感知用户实时兴趣。假设广告效果预估模型可以准确地预估展示广告所带来的期望社会福利，则选择候选广告集合中具有最大期望社会福利预测值的广告进行展示，可以最大化期望社会福利。
20. 因此，本公开提供的广告分配方法为：1. 用户对上一广告的点击及下翻行为影响预估模型对当前用户兴趣的建模，从而影响当前广告决策。2. 当进行展示广告决策时，根据广告效果预估模型实时预估值，选取具有最高期望社会福利值的广告进行展示。
21. 广告付费方法：借鉴传统的VCG机制的想法，我们设计了可以满足后验激励兼容和个体理性的付费方法，具体方式如下：
22. 步骤一：通过广告效果预估模型，预估在当前状态下，所有候选广告的点击率和期望社会福利。
23. 步骤二：根据步骤一预估的期望社会福利排序，具有最高期望社会福利的广告主记为，具有次高期望社会福利的广告主记为。假设此时预估的广告主的点击率和期望社会福利分别为
24. 步骤三：将广告主从当前状态的候选广告集合中去除，用广告效果预估模型预估广告主在去除了广告主时的期望社会福利，记为。
25. 步骤四：计算出广告主所需的点击付费。当用户点击了所推荐的广告时，广告平台向广告主收取所计算出的广告费用。
26. 可以证明本公开提出的广告付费方法满足后验激励兼容和个体理性性质，从而保证广告市场平稳运行。
27. 上述介绍了广告效果预估模型的模型结构，该模型需要通过模型训练才能实际使用。以下介绍广告效果预估模型的训练方法：
28. 我们选择双重深度Q网络（DDQN）作为我们端上重排序模型的强化学习算法。DDQN解决了动作价值的过估计问题，并将动作的选择与它们的评估分离，从而得到更稳定和准确的价值估计。此外，我们采用了经验重放技术，我们使用一个重放缓冲区来存储来自学习过程的经验。一则经验是一个由当前状态、下一个状态、动作以及动作的奖励和组成的元组，其中状态、动作、奖励均为强化学习领域中的概念，分别表示智能体面对的环境、面对该环境选择的动作以及该动作所获得的奖励。具体地，状态包括：用户资料、用户浏览历史、候选广告集合，均包括前述特征。
29. 步骤一：整理端侧设备收集的用户日志数据，通过数据清洗、整合等方式生成静态数据集。
30. 步骤二：根据静态数据集训练环境模型，此环境模型可以是任意最先进的精排模型，学习并预测用户对特定广告的点击率、下翻率。在此实施例中，我们采用了快手的端上重排模型作为环境模型。环境模型用于预测并模拟用户行为，为强化学习模型构建模拟环境。
31. 步骤三：云侧推荐系统对某用户产生精排广告集合，并提供给待训练的广告效果预测模型。
32. 步骤四：广告效果预测模型逐坑位决定展示的广告。具体来说，广告效果预测模型根据当前状态预测广告的点击率、下翻率、期望社会福利。根据预测值，以及本公开提供的广告分配及付费方法，确定当前坑位展示的广告及点击付费。
33. 环境模型则根据其预测的点击率、下翻率，模拟用户实际的点击、下翻行为，更新状态为，具体需要在用户的浏览历史中加入新浏览的广告，及候选广告集合删去刚展示的广告。
34. 将新的经验加入重放缓冲区中，重复步骤四直到广告展示完或模拟环境中用户不再下翻浏览广告。
35. 步骤五：随机抽取一批数据进行学习，其中为抽取的batch大小。根据抽取的batch构造广告效果预测模型的训练目标如下：
36. 监督学习任务：用户点击率和下翻率预测为监督学习任务，在本实施例中，我们采用了二元交叉熵作为损失函数，假设对batch中第条样本，模型预测的点击率、下翻率分别为，则定义如下公式表述的损失
37. 强化学习任务：期望社会福利需要通过强化学习方法进行学习。首先，根据双重深度Q网络算法，我们先计算出期望社会福利的目标值为：，其中，为广告的报价，分别为双重深度Q网络中的策略网络和目标网络。据此，可以根据目标期望社会福利和模型预测的期望社会福利计算最小平方误差，如下
38. 辅助任务：由于监督学习和强化学习使用了独立的结构，对这两项任务进行单独优化可能会导致预测值之间出现差距。因此，在本实施例中，我们引入了一个额外的辅助学习任务来减少这种差距
39. 其中，为状态时展示广告模型所预测的期望社会福利，为状态时展示广告模型所预测的期望社会福利。
40. 步骤七：模型最终的损失为这四个损失的加权求和。通过神经网络的反向传播最小化模型预测值与实际目标值之间的差距。



1. 步骤八：重复步骤三，直到模型收敛，学得所需的广告效果预测模型。
2. 本实施例中使用的完整训练流程如图所示：



1. 综上，本实施例中的端上实时视频流广告推荐及广告分配、广告计费的总体流程如下：
2. 步骤一：端设备向云侧服务器请求新的广告分页。具体地，首先用户开始在端上备上的一次会话，这会触发端设备向云服务器发送一个获取新分页的请求。然后，云侧的广告系统经过多个阶段，如召回、粗排和精排，并产生一个固定大小的候选广告集。这些候选广告与他们的特征随后一并发送到端设备。
3. 步骤二：决策展示给用户的视频流广告。当用户浏览完当前广告，并下翻浏览下一条广告时，预设的广告效果预估模型获取如前所述的端上实时维护的用户、广告特征，并进行处理，得到未展示的候选广告集合中各自对应的预测点击率、下翻率、期望社会福利。
4. 步骤三：按本公开提供的广告分配方法，选取具有最高打分的广告在当前广告坑位进行展示。
5. 步骤四：根据用户对当前推荐视频流广告的浏览信息，确定用户对当前广告的真实兴趣，包括以下行为：是否点击广告进入详情、是否下翻浏览下条广告。并以此更新端上维护的用户、广告特征。
6. 步骤五：如果用户点击了当前推荐的视频流广告，根据本公开提供的广告付费方法，计算广告主对该次点击的付费。
7. 步骤六：如果已耗尽云侧下发的候选广告，则重复步骤一；否则重复步骤二，直到用户不再下翻，表明用户已退出app。
8. 步骤七：将端侧记录的日志数据回传至云侧，供环境模型及广告效果预测模型训练。

**模拟实验结果**

1. 我们在Mobile的公开数据集上进行了模拟实验。Mobile是一个针对栏目推荐场景的数据集。它记录了用户连续八天的会话数据，包含了24.6万个会话、1.2万次点击、3.5万名用户和8个栏目，其中栏目数量少是栏目推荐场景的一个特点。因此，在这个数据集中广告位的数量为8个。我们采用随机生成方法来创建缺失的竞价数据。具体来说，我们假设广告商的出价遵循一个正态分布，其中且方差遵循均匀分布。通过调整参数，我们可以控制广告商出价的分布。
2. 我们设置了以下三个基线广告分配与付费方法，与我们提出的方法进行对比：1.贪婪GSP：根据广告的千次广告收入（eCPM）排序，采用广义二价付费（GSP）；2.uGSP：采用了和本公开相同的分配方法，但是计费方法不同；3.AdaAuc-2:在本公开提出的分配方法中，展示广告对未来期望社会福利的影响权重调整为两倍。
3. 我们统计了不同方法下，达到的放缩期望社会福利（SSW）、会话长度（SL）和点击次数（NC），分别反映了不同方法最大化社会福利的能力，以及满足用户需求的能力。具体结果如图2和图3。
4. 图2中展示了我们提出的分配与计费方法可以达到最高的放缩期望社会福利。因为AdaAuc-2在分配广告时对未来期望社会福利的影响权重调整为了两倍，因此它可以达到更长的会话长度，以及得到更多的点击。但AdaAuc-2并没有因此获得更高的放缩期望社会福利，这反映了AdaAuc-2高估了广告对未来期望社会福利的影响，而AdaAuc很好地平衡了当前与未来收益的关系。
5. 此外，我们研究了不准确的预估模型对后验激励兼容性质的影响。在我们的实验中，我们随机选择一个广告主，并将其出价设置为其真实估值的𝛼倍，其中𝛼∈[0, 2]，同时所有其他广告商都诚实出价。然后，我们分析在不同出价策略𝛼下不诚实广告商的平均效用。如图4所示，Greedy-GSP、uGSP、AdaAuc-2和AdaAuc都存在违反后验IC性质的问题，因为广告商获得最大效用的点不在诚实出价点𝛼=1上。然而，在这些方法中，使用AdaAuc时，达到广告商最优效用的策略最接近诚实出价策略，这表明AdaAuc在后验IC性质上表现更好。此外，AdaAuc的曲线比其他基线方法的曲线更陡，这意味着AdaAuc对不诚实出价行为施加了更严重的惩罚。总之，尽管不准确的广告效果预估模型损害了AdaAuc的后验激励兼容性质，但实验证据表明，与Greedy-GSP相比，AdaAuc面对具有策略性的广告主时受到的影响较小。因此，它可以更好地激励广告商提交真实的出价。此外， AdaAuc的性质有数学保障，因此提出更准确的模型可以保证后验IC。

**说 明 书 附 图**

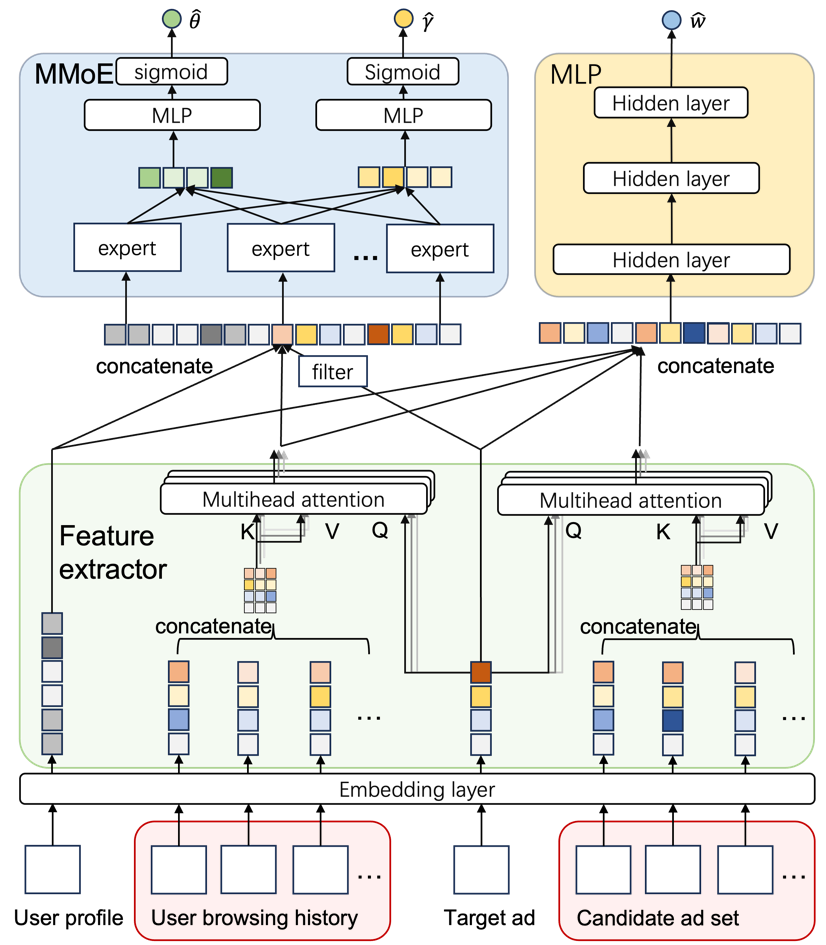


图 1

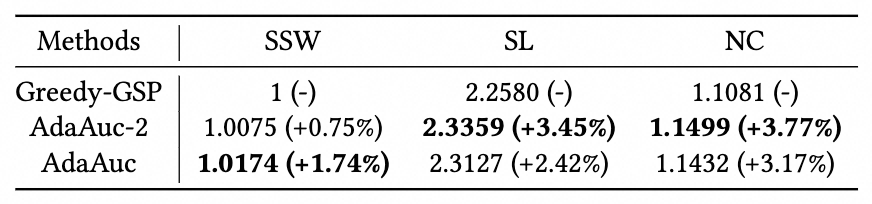


图 2

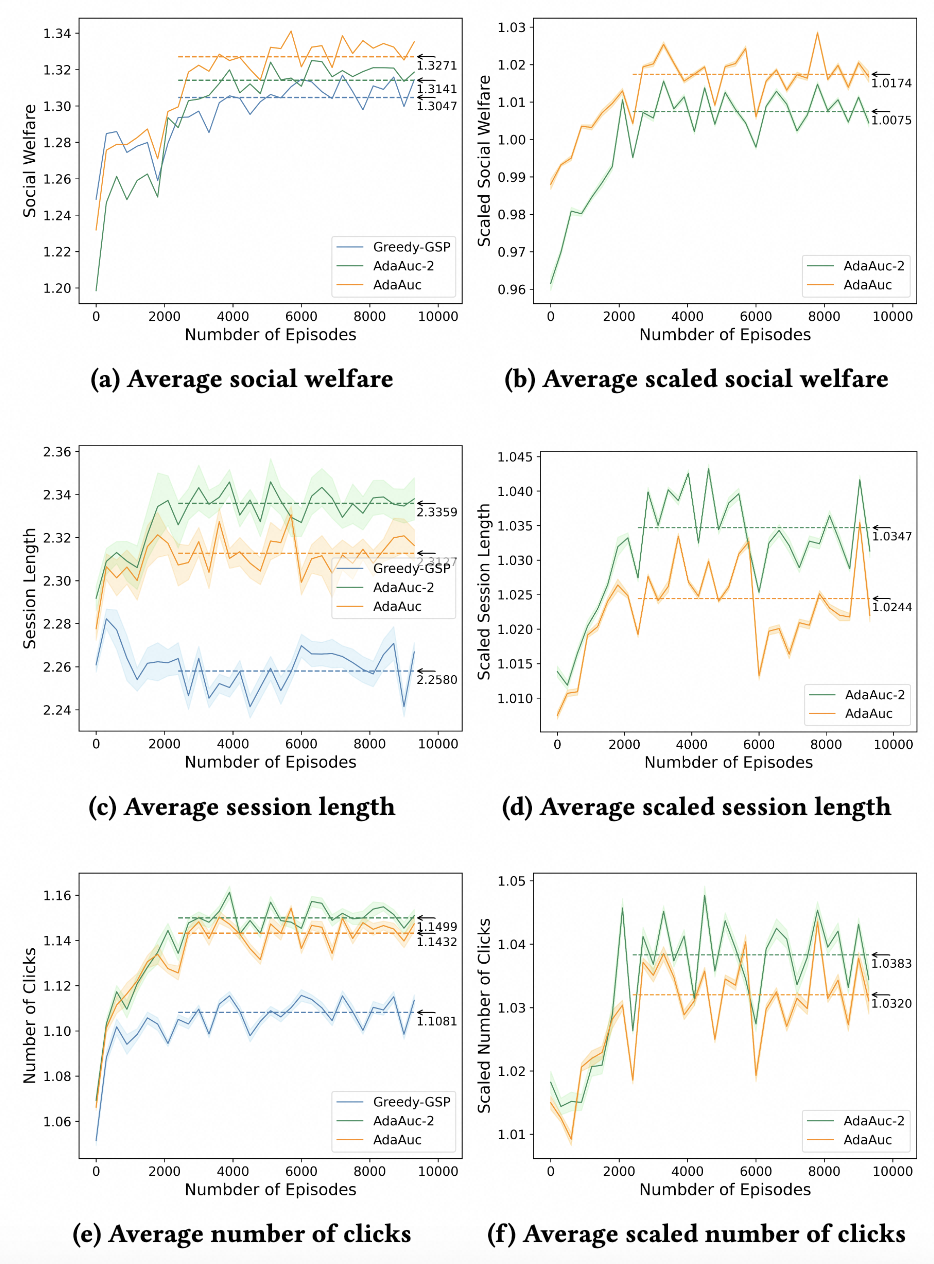


图 3

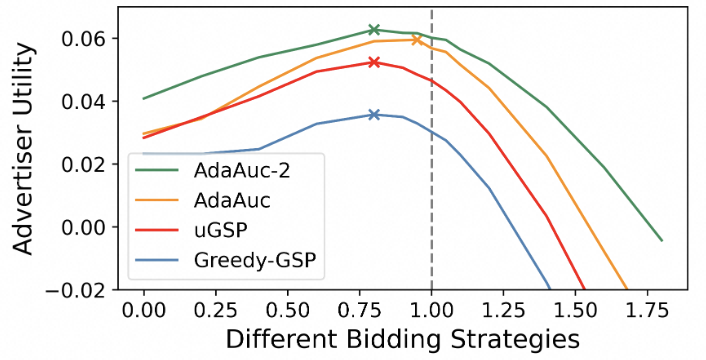


图 4

**权 利 要 求 书**

1. 一

**说 明 书 摘 要**

本公开提供了一种视频流广告的投放效果的预测方法，还提供了一种最大化平台利益的广告分配及计费方法。预测方法包括：基于强化学习的端上重排序模型，对用户端上实时行为进行建模，并得到预估广告的点击概率、下翻概率及带来的期望社会福利。广告分配方法包括：基于所预测的不同广告的点击概率、下翻概率及带来的期望社会福利，选出并展示最大化社会福利的广告。广告计费方法包括：基于所预测的不同广告的点击概率、下翻概率及带来的期望社会福利，计算广告对社会福利的边际贡献作为广告的付费。由此，通过对用户实时行为建模，本公开实施例的方案可以有效提升对用户点击、下翻行为概率进行预估的准确性，并优化用户体验；通过设计的广告分配方法可以显著提升平台长期收益；通过设计的广告计费方法，本公开实施例的方案可以保障广告市场的稳定性，激励广告主参与并投放广告，从而提升平台收益。

论文摘要部分：重排序模型已经广泛应用于移动设备中的推荐系统和在线广告。这些模型以用户在设备上的反馈作为输入，并根据用户实时兴趣动态地重新排序项目。然而，基于预测值贪婪地重排序项目而不考虑其对未来的影响，即项目的外部性，可能会导致预期奖励的次优。此外，直接采用常用的广义第二价格（GSP）和Vickrey-Clarke-Groves（VCG）拍卖的支付规则不能保证经济属性，并可能导致在线广告市场的不稳定。在本文中，我们提出了一种适应性广告拍卖方法，称为AdaAuc，用于移动设备上的广告系统。首先，我们将流媒体广告中的设备上广告重排序建模为马尔可夫决策过程（MDP），并设计了一种新的分配规则来优化预期的社会福利，考虑了广告的外部性。我们提出了一种类VCG的支付规则，该规则在MDP框架内动态定价显示的广告。我们证明AdaAuc满足事后激励兼容性（ex-post IC）和个体理性（IR）两个经济属性，因此保证了广告商的真实出价。最后，我们设计了一个端到端的深度强化学习（RL）框架来训练重排序模型，它适用于资源有限的设备，以实现实时的设备上拍卖。在公共数据集上的实验结果证明了AdaAuc在预期社会福利方面的有效性。