**说 明 书**

**一种考虑不确定状态的自动出价智能体**

**技术领域**

1. 本发明涉及的是一种电商广告拍卖中的实时竞价（Real-time Bidding）方法，具体是一种在广告拍卖过程中考虑特征不确定性的出价策略调控方法，在进行序列化决策时利用特征分布与强化学习方法，构建实时自动竞价智能体，以较低的复杂度成本，显著的提高了智能体实时竞价所得到的转化量（例如该商品被购买的数量）。

**背景技术**

1. 受限于互联网平台的海量流量和广告主有限的手动调控能力，在电商广告中，为了更好的满足广告主优化转化量的诉求，平台会为广告主提供自动出价智能体，协助广告主高频调控出价策略，以获得更多的转化量。自动出价智能体的效率和优化问题非常重要。
2. 构建自动出价智能体的问题，其任务是在未来环境未知的前提下，根据观测到的反馈信息，实时调整出价策略来提升最终转化量。近年来，很多基于该思想的自动出价方法被提出。这些构建方法大都假设出价智能体能够实时观测到真实反馈，从而合理的调整实时出价策略。但是，现有的出价策略调整智能体，都忽略了真实环境中的反馈延迟性带来的特征不确定性，不论这些策略调控方法的理论效果如何，若其无法获得真实的反馈，决策的效果便会大打折扣。
3. 反馈特征不确定性，是线上真实环境区别于离线理想环境的一条重要性质。它是由用户心智的复杂性和用户行为随机性所决定的。在电商广告环境中，当用户浏览并点击了一条广告后，其是否会发生转化行为，转化行为何时发生，都具有较大的不确定性。转化行为的延迟可长达几小时甚至几天，这对具有高反馈实时性要求的自动出价智能体的决策效果有较大的影响，会带来优化效果的损失，并且增加了超限风险。
4. 所述的超限风险是指广告主会为广告计划设置单位转化成本，即竞得一次转化所需的花费。而因为种种原因，自动出价智能体的竞拍结果导致真实的单位转化成本超过广告主设置的预期值，则被称为超限现象。超限现象意味着自动出价智能体没有很好的完成目标任务，是其设计中应当考虑的重点。
5. 经过对现有技术的检索发现，Yue He等人在《A Unified Solution to Constrained Bidding in Online Display Advertising》（KDD’2020）中比较和总结了以往的自动出价自动出价智能体构建方式，将带约束的实时出价问题构建为线性规划问题，并通过求解推导，将出价问题解耦为出价策略与流量质量两个方面，而自动出价智能体只需要考虑调控出价策略，降低了问题的整体复杂度。其提出了一种面对多约束环境的通用出价智能体构建方法USCB。USCB基于强化学习（Reinforcement Learning）结构，能够基于实时反馈构建当前的状态（State），并根据当前状态快速调控出价策略，取得了良好的优化效果。但是该技术仍然假设反馈具有即时性，没有考虑到反馈延迟与状态的不确定性，在真实的线上环境中的存在偏差与优化效果的损失。

**发明内容**

1. 本发明针对现有技术存在的上述不足，提出一种参考不确定状态的出价决策智能体。首先利用已有的先后验信息构建单位转化成本的分布，使得当前智能体对当前状况有更清晰的预期；然后构建能参考该分布进行实时决策的智能体，提升整体的转化量数量，降低超限风险。
2. 本发明是通过以下技术方案实现的，本方法包括以下步骤：
3. 步骤1、结合前链路传递的先验转化信息和智能体真实观测的后验转化信息，利用转化延迟分布模型，参考其他即时反馈信息，得到当前智能体竞得流量的单位转化成本的分布。
4. 所述的前链路传递的先验转化信息是指：电商广告中，在智能体出价之前，平台会提供给其某一条流量的预估转化率以供参考。
5. 所述的真实观测的后验转化信息是指：在智能体竞得一条流量后，在某一时刻观测该流量转化结果，若可以观测到流量转化，称为正后验信息；尚未观测到该流量转化，称为负后验信息。
6. 所述的转化延迟分布模型是指：在流量最终转化的前提下，流量转化延迟的分布。表示为：，其中表示从流量点击到当前观测所经过的时间。该分布的物理意义为若流量最终发生转化，其转化延迟小于的概率。
7. 所述的其他即时反馈信息是指：智能体的实时花费，出价周期剩余的时间等信息，这些信息具有即时反馈性和确定性。
8. 步骤2、对自动出价问题进行形式化建模，并利用强化学习方式来解确定状态下的该问题。
9. 所述的形式化建模指是指：把离线环境（所有流量属性已知）中的出价决策问题建模成线性规划问题：
10. 优化目标：
11. 限制条件：，其中：表示竞拍周期中的流量总数，表示是否选择竞得某条流量，表示流量的价值，表示竞得流量的花销，表示广告主的预算，表示广告主预设的目标单位转化成本。
12. 所述的线性规划问题的物理意义为，在存在预算约束与单位转化成本约束的前提下，智能体尽量多的选取高性价比的流量。
13. 所述的线性规划问题在离线环境中存在最优解，最优出价形式为bid = ，其中和为线性规划问题对偶问题中，两个限制条件对应的对偶因子。
14. 所述的离线环境是指利用往期数据构建的模拟环境，其中智能体能够获知全天内的所有流量的价值信息
15. 所述的强化学习方式是指：通常情况下，智能体实际上没有能力观测到全天的所有流量信息（如智能体在中午时无法获知晚上流量的具体质量），所以上述的线性规划建模无法直接应用到工业环境中，需要将该问题建模为马尔科夫决策过程（MDP）问题。再使用强化学习方式近似拟合该决策过程。强化学习结构包括策略（Policy）网络和动作状态函数（Q value）网络。
16. 所述的马尔科夫决策过程是指，对于时刻的状态，其可能的分布仅与时刻的状态及智能体的决策动作有关，而与时刻之前的所有状态或智能体决策动作均无关。
17. 所述的时刻的状态是指智能体对于当前环境与自身性质的认知。
18. 所述的策略网络是指：输入为智能体当前所处状态，输出为智能体策略动作的神经网络。当智能体获知当前的状态时，其参考该网络确定决策动作。
19. 所述的动作状态函数网络是指：输入为智能体当前所处状态与智能体的决策动作，输出为动作状态函数值的网络。当智能体获得当前的状态时，其可以获知每一种决策所对应的剩余时间收益和，表示为。
20. 步骤3、考虑当前状态的不确定性，参考当前状态的离散分布，利用不确定状态理论，结合强化学习模型中的动作状态函数网络，构建自动出价决策智能体。
21. 所述的状态不确定性是指：在真实环境中，某些物理量的反馈存在延迟（如单位转化成本信息），出价智能体无法获取当前的准确状态。出价智能体依赖实时状态进行决策。可以使用对应的分布信息表达该物理量的不确定性。
22. 所述的离散分布是指：对于有限的个可能的状态，智能体可以获知其当前所处特定状态，的概率，。
23. 所述的不确定状态理论是指：，其中表示当前最优的出决策，表示可能状态的离散分布。其物理意义为，最优的决策应当能够对所有可能状态负责。

**技术效果**

1. 与现有技术相比，本发明优点包括：可以在存在延迟反馈的环境中，显著提高出价智能体的优化效果，并降低超限现象（指最终的单位转化成本大于广告主预设的值）发生的概率；实现了离线训练和线上推理过程的解耦，引入的额外计算代价非常低；实现简单，可以部署到现有的强化学习决策框架中。

**附图说明**

1. 图1为模拟实验中重复单一广告计划下，GQOUS算法与现有算法的对比。
2. 图2为模拟实验中不同广告计划下，GQOUS算法与现有算法的对比。
3. 图3为模拟实验的具体竞得转化次数对比示意图。

**具体实施方式**

1. 下面对本发明的实施例作详细说明，本实施例在以本发明技术方案为前提下进行实施，给出了详细的实施方式和具体的操作过程，但本发明的保护范围不限于下述的实施例。

**实施例1**

1. 本实施例包括以下步骤：
2. 第一步、结合前后验信息构建单位转化成本的分布：自动出价智能体需要参考实时转化量和单位转化成本信息进行决策。在智能体对于已竞得流量的最终转化量的估计过程中，有效信息包括先验信息和后验信息。先验信息为预估转化率，相对确定，而后验信息会随着时间发展不断演化。
3. 对于已经发生点击的流量，引入事件表示智能体在时刻观测流量，其是否已经发生转化，表示智能体观测到转化，而表示尚未观测到转化；引入事件表示流量最终是否发生转化，表示流量最终发生转化，而表示最终未发生转化。
4. 经过上述定义后，对于流量在时刻的转化总量的预估被重新定义为
5. 所述的表示后验观测条件下流量最终转化的概率。
6. 若已经观测到流量转化，显然流量必然转化。即，则可以推出。将这部分流量的总量记作 ；
7. 而对于未观测到转化的流量，经过贝叶斯变分过程，能够获得

所述的是指流量最终转化的条件下，智能体尚未观测到其转化的概率，其物理意义等价于：该流量最终转化的条件下，，其中为该流量的转化延迟，为从点击发生到当前观测所经过的时间；而即为流量发生转化的先验概率，为。

1. 在上述物理意义的等价下，可以计算得：
2. 经过上述处理后，可以得到时刻转化总量的无偏估计：
3. 所述的时刻转化总量的无偏估计形式下，每一条流量都相互独立，且为伯努利变量形式。
4. 所述的伯努利变量形式指流量会以概率取1，表示流量发生转化；会以概率(1-取0，表示流量未转化。
5. 经过上述建模后，转化量总量表达为异质独立伯努利变量的累加和形式，可以通过中心极限定理还原为分布形式，该分布为最终转化总量可能结果的分布。
6. 所述的转化总量分布呈现出高斯分布的形式，表达了智能体对于环境不确定性的认知。
7. 单位转化成本的分布由实时花费除以上述转化量分布得到。
8. 第二步、把参考单位转化成本的出价过程建模成不确定状态下的连续决策问题
9. 经过上述建模后则将在线出价问题重新建模为存在状态不确定性的序列决策问题。可以据此设计考虑输入状态为分布形式的自动出价智能体。本实例中包括序列化决策算法GQOUS。
10. 所述的GQOUS算法是指：根据优化问题的约束条件，使用强化学习的方式构建确定性环境下的出价决策模型：该模型将当前所处状态作为参考，进行出价决策。从上一步中所述的该算法具体步骤包括：
11. i) 初始化模型参数，使用往期样本，令模型在模拟环境中不断与环境进行交互，进行确定性模型的训练。设置惩罚项来进行模型行为调控。获得确定性条件下的决策模型和对应的动作状态函数模型，留作备用。实时单位转化成本为动作状态函数模型的重要参数。
12. ii) 在线的不确定环境中，自动出价智能体观测到当前的状态分布信息。智能体从该连续分布中进行分位数采样，获得该状态分布对应的离散分布。
13. iii) 智能体参考离散分布进行决策，对于离散分布中个可能的状态，，设置个动作状态函数模型，他们的输入均留空。将个动作状态函数模型并联起来，他们共享相同的决策动作。对于每一个输入，将所有并联网络的输出结果相加得到，即为考虑离散状态分布情况下的全局动作状态函数。
14. iv) 获得能使全局动作状态函数取得最大值决策动作，则为本次决策过程中的最优解。
15. 所述的往期样本指某一天内的流量记录，其中包括每条流量的预估点击率、预估转化率、赢得该流量所需的出价等。
16. 所述的模拟环境指利用一天内的往期样本，模拟智能体进行在线出价竞争流量的过程。
17. 所述的惩罚项指若智能体在模拟环境中模拟结果出现了超限现象，应当对其获得的收益进行相应削减，表示超限控制的目的。
18. 所述的确定性条件指不存在反馈延迟的理想条件，在该条件下智能体可以准确的获得其决策依赖的所有信息。

**模拟实验结果**

1. 本实施例的模拟试验是在一个模仿线上环境的仿真竞价场景中进行的，该场景提供了一天内的百万量级流量信息，包含每条流量的时间、预估点击率、预估转化率以及竞得该流量所需的出价等信息。在试验中，为了贴合线上环境的转化行为特征，为其中模拟的转化赋予了稀疏性和延迟性。设置USCB算法的policy模型和Q-value模型作为对比项，构建了两个实验对比效果：实验一：统计1000个不同场景的广告计划的平均效果，每个广告计划运行一次；实验二：重复同一广告计划1000次，考察1000次的平均效果，共取5个计划做对比。
2. 所述实验中的稀疏性是指对模拟竞得的流量的点击和转化结果通过0-1采样获得，此阿阳的参数是对应的预估点击率和预估转化率；
3. 所述的实验中的延迟性是指对于判定为转化的流量，根据延迟分布随机赋予其转化延迟，只有持有该流量超过该延迟才能观测到其转化。
4. 如图1所示，为模拟实验的5个广告计划重复1000次平均结果的比较。对比了USCB算法的策略网络、动作状态函数网络与GQOUS算法的效果。其中参数为相对转化量值，为超限率值。从图中可以看出，GQOUS算法构建的出价智能体在单广告计划的平均意义上具有更高转化量收益，且能够有效降低计划超限率。
5. 如图2所示，为模拟实验的1000个不同广告计划平均效果的比较。其中参数为相对转化量值，为超限率值。从图中可以看出，GQOUS算法构建的智能体在不同广告计划上提高了在天级别的竞得转化量结果，并且显著降低了计划超限率。从中可以得出结论，具有较强的范化性和通用性。
6. 如图3所示，为GQOUS算法和其他两种算法在1000次重复实验中的最终转化量结果统计。从图中可以看出，GQOUS算法的竞得转化量的数量普遍较高，并且其最终结果具有更高的稳定性。从中可以得出结论，GQOUS的优化效果更高，且鲁棒性优秀。

**说 明 书 附 图**

图1

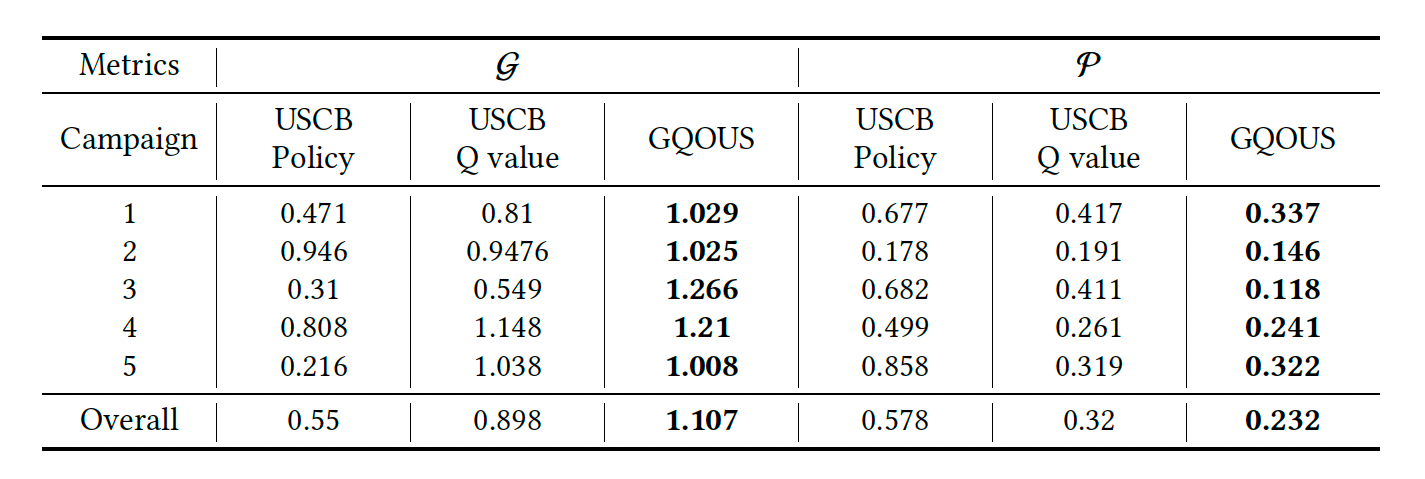


图2

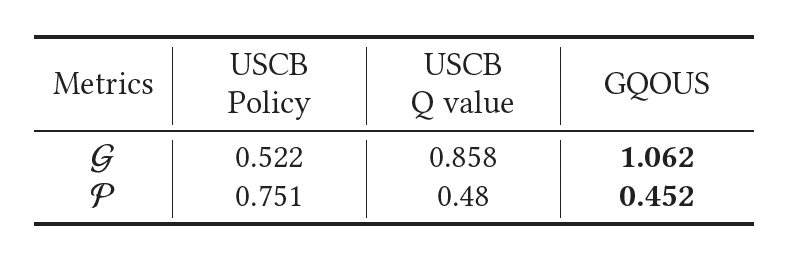
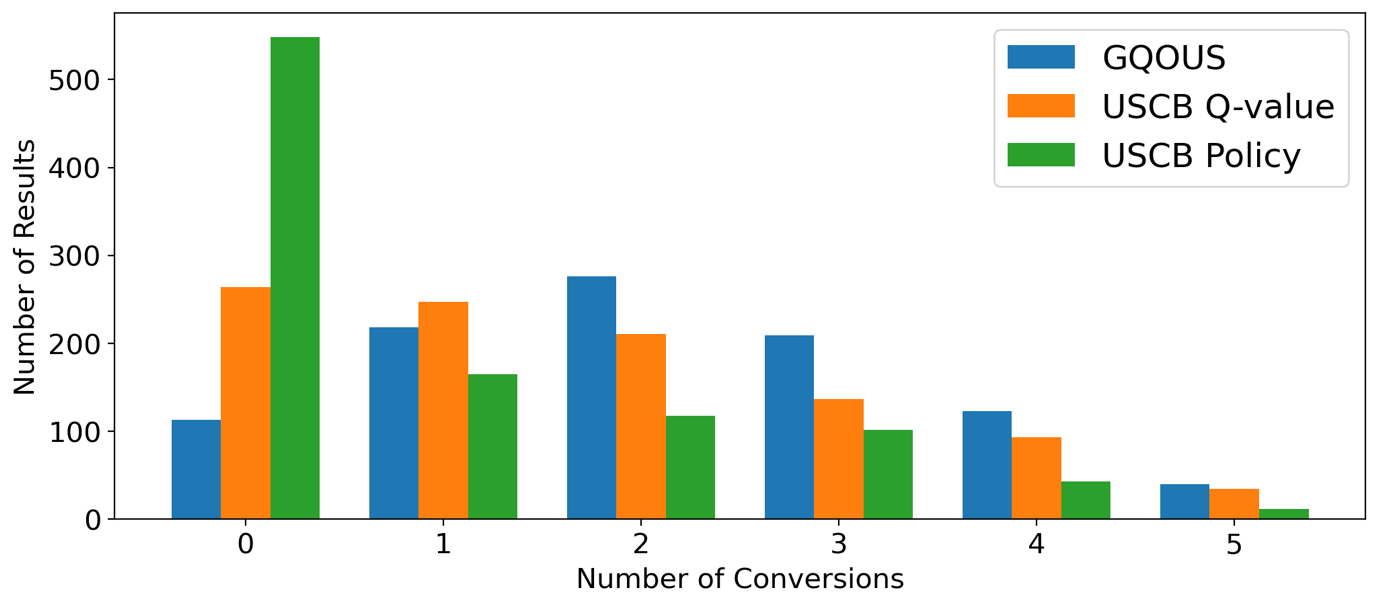


图3



**权 利 要 求 书**

**说 明 书 摘 要**

**摘 要 附 图**