

权利要求书

1. 一种利用深度强化学习进行推荐营销的方法，所述方法包括：
获取执行环境的状态信息，所述状态信息至少包括当前用户的用户信息；
根据营销策略，针对所述状态信息确定相应的营销行为，所述营销行为包括营销渠道、营销内容和营销时段的组合；
获取执行环境针对所述营销行为的当前奖励分数；
至少根据所述当前奖励分数更新所述营销策略。
2. 根据权利要求 1 所述的方法，其中所述用户信息包括，用户基本属性信息，以及用户历史行为信息。
- 10 3. 根据权利要求 2 所述的方法，其中所述用户历史行为信息包括以下中的至少一项：历史操作序列、最近浏览页面列表、最近接受的营销信息列表、消费记录、交易记录、贷款记录、出行记录、保险记录。
4. 根据权利要求 1 所述的方法，其中所述用户信息包括用户画像信息，所述用户画像信息基于用户基本属性信息和/或用户历史行为信息预先确定。
- 15 5. 根据权利要求 1 所述的方法，其中所述状态信息还包括，所述当前用户所处场景的场景信息。
6. 根据权利要求 5 所述的方法，其中所述场景信息包括，位置服务场景信息，支付场景信息，钱包操作事件信息。
7. 根据权利要求 1 所述的方法，其中所述营销内容包括营销消息和利益包，所述利益包包括以下中的一种或多种：红包、优惠券、权益。
- 20 8. 根据权利要求 1 所述的方法，其中所述营销渠道包括以下中的一种或多种：消息推送，卡片展示，角标提醒，红点提醒。
9. 根据权利要求 1 所述的方法，还包括，至少基于所述当前奖励分数确定回报价值，所述回报价值包括所述当前奖励分数和至少一项未来奖励分数的加权求和；
- 25 所述至少根据所述当前奖励分数更新所述营销策略包括，确定后续营销策略，使得所述回报价值最大化。
10. 一种利用深度强化学习进行推荐营销的装置，所述装置包括：
状态获取单元，配置为获取执行环境的状态信息，所述状态信息至少包括当前用户的用户信息；
行为确定单元，配置为根据营销策略，针对所述状态信息确定相应的营销行为，所述营销行为包括营销渠道、营销内容和营销时段的组合；
- 30

分数获取单元，配置为获取执行环境针对所述营销行为的当前奖励分数；

更新单元，配置为至少根据所述当前奖励分数更新所述营销策略。

11. 根据权利要求 10 所述的装置，其中所述用户信息包括，用户基本属性信息，以及用户历史行为信息。

5 12. 根据权利要求 11 所述的装置，其中所述用户历史行为信息包括以下中的至少一项：历史操作序列、最近浏览页面列表、最近接受的营销信息列表、消费记录、交易记录、贷款记录、出行记录、保险记录。

13. 根据权利要求 10 所述的装置，其中所述用户信息包括用户画像信息，所述用户画像信息基于用户基本属性信息和/或用户历史行为信息预先确定。

10 14. 根据权利要求 10 所述的装置，其中所述状态信息还包括，所述当前用户所处场景的场景信息。

15. 根据权利要求 14 所述的装置，其中所述场景信息包括，位置服务场景信息，支付场景信息，钱包操作事件信息。

15 16. 根据权利要求 10 所述的装置，其中所述营销内容包括营销消息和利益包，所述利益包包括以下中的一种或多种：红包、优惠券、权益。

17. 根据权利要求 10 所述的装置，其中所述营销渠道包括以下中的一种或多种：消息推送，卡片展示，角标提醒，红点提醒。

20 18. 根据权利要求 10 所述的装置，还包括回报价值确定单元，配置为至少基于所述当前奖励分数确定回报价值，所述回报价值包括所述当前奖励分数和至少一项未来奖励分数的加权求和；

所述更新单元配置为，确定后续营销策略，使得所述回报价值最大化。

19. 一种计算设备，包括存储器和处理器，其特征在于，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现权利要求 1-7 中任一项所述的方法。