本发明属于计算机网络拥塞控制领域，公开了一种结合强化学习和ECN显式拥塞通知机制，对BBR v1拥塞控制协议的动态参数优化方法。传统BBR协议中，关键参数固定、难以适应复杂的网络环境。本发明基于BBR状态机，使用强化学习算法动态生成对BBR关键参数，包括拥塞窗口增益、最小RTT探测周期的调整策略，使协议能够根据当前的链路状态，对关键参数进行动态优化。同时在BBR中集成ECN显式拥塞通知机制，允许客户端向服务器反馈拥塞信息，增强协议对链路状态的实时感知能力和快速拥塞响应能力。该方法适用于数据中心网络、无线网络等多种场景，纯软件实现，不依赖特定的硬件，易于在网络协议栈中部署。基于网络模拟器ns3的测试表明，该方法能够有效的提高网络传输性能，增大吞吐量的同时，降低往返时延和丢包率。

1.一种基于BBR v1拥塞控制协议的动态参数优化系统，包括发送端和接收端，其特征在于：发送端包括强化学习模型和网络协议栈，所述强化学习模型可以根据发送端网络协议栈实时采集链路状态特征，动态生成对BBR关键参数，包括拥塞窗口增益、最小RTT探测周期的调整策略。所述网络协议栈包括BBR状态机，该BBR状态机根据所述调整策略，优化BBR关键参数。接收端包括拥塞信息反馈模块，接收端收到ECN数据包时，将拥塞信息反馈给发送端，进而对发送端的拥塞窗口增益、最小RTT探测周期进行细粒度调整，同时作为发送端的所述强化学习模型的输入之一，协助生成所述参数调整策略。

2.根据权利要求1所述的系统，其特征在于：所述接收端的拥塞信息反馈模块，接收端收到携带CE拥塞经历标记的数据包后，将拥塞信息，即携带ECN-Echo标记的ACK反馈给发送端。

3.根据权利要求1所述的系统，其特征在于：发送端收到拥塞信息，即携带ECN-Echo标记的ACK后，使用该拥塞信息对拥塞窗口增益、最小RTT探测周期进行细粒度调整，细粒度调整的公式为：

其中，是一个RTT内，发送端收到的ACK中，携带ECN-Echo标记的ACK数量占ACK总数量的比例；是拥塞窗口增益；是最小RTT探测周期；和是BBR持续维护的瓶颈带宽和最小RTT；是拥塞窗口大小。

4.根据据权利要求1所述的系统，其特征在于：发送端会持续维护指数加权移动平均的RTT抖动，用于响应链路状态的突发改变，指数加权移动平均的RTT抖动的计算公式为：

其中，是指数加权移动平均的RTT抖动；是平滑因子；是当前的RTT测量值。

5.根据权利要求1所述的系统，其特征在于：所述强化学习模型采用PPO近端策略优化的强化学习算法，结合离线学习和在线学习对模型进行训练和更新，更新过程用公式表示为：

、、、和表示吞吐奖励权重、时延惩罚权重、拥塞惩罚权重和公平奖励权重系数。可以适当增大的权重，使得算法倾向于探索高带宽利用率；也可以适当减小的权重，降低算法对时延的敏感度，以容忍一定程度的队列堆积；也可以适当减小的权重，降低算法对拥塞的敏感度，以容忍短暂的网络拥塞；通过维持基础公平奖励权重，保证多流环境下的公平性。是当前的吞吐量，即最近一个RTT交付的数据量；是近期10个RTT内的平均丢包率；是Jain公平性指数。

6.根据权利要求1所述的系统，其特征在于：所述强化学习模型可以适用于不同的网络环境，例如数据中心网络或无线网络。通过实时采集链路特征，动态优化BBR关键参数，保证在不同网络环境下提高网络传输性能。

7.一种对BBR关键参数的动态优化方法，其特征在于：该方法分为以下步骤：

步骤1、发送端和接收端运行BBR协议，发送端运行前可以加载对应网络环境下的预训练参数，也可以使用默认参数；

步骤2、运行时发送端网络协议栈中的强化学习模型根据实时采集的链路状态特征，动态生成对BBR关键参数，包括拥塞窗口增益、最小RTT探测周期的调整策略；

步骤3、将步骤2中生成的调整策略应用于BBR协议，更新BBR关键参数；

步骤4、接收端收到携带CE拥塞经历标记的数据包后，将拥塞信息，即携带ECN-Echo标记的ACK反馈给发送端；

步骤5、发送端收到携带ECN-Echo标记的ACK后，细粒度调整拥塞窗口增益、最小RTT探测周期；同时作为发送端的所述强化学习模型的输入之一，协助生成所述参数调整策略。

步骤6、发送端使用优化后的BBR关键参数，继续发送后续数据流。

8.根据权利要求7所述的，对BBR关键参数的动态优化方法，其特征在于：步骤1中实时采集的链路状态特征，包括链路瓶颈带宽和最小RTT、当前的交付速率、近期的RTT抖动和丢包率等，以及所述接收端反馈的拥塞信息，一并作为所述发送端的强化学习模型的输入，以动态生成所述参数调整策略。

**基于强化学习和ECN显式拥塞通知机制的BBR协议优化方法**

**技术领域**

本发明属于计算机网络拥塞控制领域，特别涉及一种结合强化学习和ECN显式拥塞通知机制，优化BBR v1协议。具体的，本发明提出了一种在BBR中集成ECN显式拥塞通知机制，使用强化学习算法动态优化BBR关键参数的方法，包括拥塞窗口增益、最小RTT探测周期。纯软件实现，不依赖特定的硬件。适用于数据中心网络、无线网络，以及其他复杂的网络环境，能够在高动态的链路环境下实时调整拥塞控制策略，提升吞吐量、降低往返时延和丢包率，提高网络传输性能。

**背景技术**

与当前广泛使用的拥塞控制算法CUBIC协议不同，BBR v1协议不再以丢包作为唯一拥塞信号，而是通过持续维护最大交付速率作为链路的瓶颈带宽，和队列0积压时的最小RTT，以调整拥塞窗口。对比传统的以丢包作为拥塞信号的算法，BBR没有传统的慢启动阶段，并且由于BBR主动探测链路状态，对比传统的被动拥塞避免的算法，可以提高吞吐量，减少队列堆积，进而降低数据包排队时延。然而BBR也存在明显缺陷。BBR有时存在较高的带宽估计，倾向于抢占CUBIC等算法的带宽，导致多流环境下BBR的公平性较差；并且，BBR在数据中心网络，无线网络等网络环境下表现较差。主要的，BBR算法的关键参数，例如拥塞窗口增益、最小RTT探测周期都是静态配置，拥塞窗口增益固定为2，最小RTT探测周期固定为10秒，对于链路状态突发改变的响应相对滞后。网络突发拥塞时，BBR难以及时收敛到合适的发送速率，缺少针对实时链路状态的动态参数优化方法。

DCTCP协议是为数据中心网络设计的拥塞控制算法，支持ECN的交换机会实时测量瞬时队列长度，当瞬时队列长度超过阈值时，交换机会在数据包中设置CE标记，表示拥塞经历。接收端收到携带CE标记的数据包后，通过在ACK中设置ECN-Echo标记，将拥塞信息反馈给发送端。对比传统的拥塞控制算法，DCTCP在平滑处理后，对拥塞窗口进行精细调整，而非简单地加倍或减半，以快速响应链路状态的突发改变。然而DCTCP不具备主动探测链路瓶颈带宽的机制，限制了DCTCP在复杂网络环境中的表现。

综上所述，现有拥塞控制算法在面对高动态的网络环境时存在诸多不足，通过在发送端的BBR状态机中嵌入强化学习模型，结合离线学习和在线学习对模型进行训练和更新，发送端发送前可以使用对应网络环境下的预训练参数，发送时可以根据实时链路状态在线学习并动态生成参数调整策略。同时计算每个RTT内收到的ACK中，携带ECN-Echo标记的ACK数量占ACK总数量的比例，该比例一方面对拥塞窗口增益、最小RTT探测周期进行细粒度调整，另一方面作为强化学习模型的输入之一，进一步增强BBR对链路状态的实时感知能力和快速拥塞响应能力。

**发明内容**

本发明提出了一种对BBR v1拥塞控制协议的改进方法，旨在解决BBR关键参数固定，对动态网络环境适应性较差的问题，通过结合强化学习和ECN显式拥塞通知机制，动态优化BBR关键参数，包括拥塞窗口增益、最小RTT探测周期。特别适用于数据中心网络和无线网络等高动态的网络环境，能够提高网络传输性能，增大吞吐量的同时，降低往返时延和丢包率。

为了达到上述目的，本发明是通过以下技术方案实现的：

步骤1、构建BBR状态机，作为后续强化学习模型的基础调度框架。BBR状态机包括四个阶段：Startup启动阶段、Drain队列排空阶段、ProbeBW带宽探测阶段、ProbeRTT最小RTT探测阶段。对于每个阶段，BBR有不同的链路状态探测与拥塞控制策略，后续强化学习模型作用于BBR状态机的带宽探测阶段和最小RTT探测阶段。

步骤2、发送端在数据传输过程中，实时探测当前的链路状态特征，包括链路瓶颈带宽和最小RTT、当前的交付速率、近期的RTT抖动和丢包率等，输入到发送端的强化学习模型。

步骤3、BBR协议中集成ECN显式拥塞通知机制。具体的，在运行BBR协议的接收端添加对携带CE标记的数据包的处理逻辑，发送端添加对携带ECN-Echo标记的ACK的处理逻辑。计算每个RTT内收到的ACK中，携带ECN-Echo标记的ACK数量占ACK总数量的比例，以对发送端的拥塞窗口增益、最小RTT探测周期进行细粒度调整，同时作为发送端的强化学习模型的输入之一。

步骤4、在发送端的BBR状态机中，嵌入强化学习模型。将实时采集的链路状态特征和ECN-Echo标记比例一并输入到强化学习模型，计算奖励值，动态生成对BBR关键参数，包括拥塞窗口增益、最小RTT探测周期的调整策略。

步骤5、结合离线学习和在线学习，对强化学习模型进行训练，离线学习用于预训练，在线学习用于微调BBR关键参数。

本发明的进一步改进在于：在所述步骤2中，发送端会持续维护指数加权移动平均的RTT抖动，用于响应链路状态的突发改变，指数加权移动平均的RTT抖动的计算公式为：

其中，是指数加权移动平均的RTT抖动；是平滑因子；是当前的RTT测量值；是BBR持续维护的最小RTT。本发明实施例中，取0.0625。

本发明的进一步改进在于：在所述步骤3中，发送端收到携带ECN-Echo标记的ACK时，动态更新拥塞窗口增益和最小RTT探测周期，公式为：

其中，是一个RTT内，发送端收到的ACK中，携带ECN-Echo标记的ACK数量占ACK总数量的比例；是拥塞窗口增益；是最小RTT探测周期；和是BBR持续维护的瓶颈带宽和最小RTT；是拥塞窗口大小。本发明实施例中，的初始值取2，的初始值取15秒。

本发明的进一步改进在于：在所述步骤4中，状态空间设计为：

其中，是当前的交付速率，通过计算当前RTT内，接收端确认的字节数得到；是近期10个RTT内的平均丢包率。

动作空间设计为：

其中，是拥塞窗口增益修正量；是最小RTT探测周期调整量。本发明实施例中，的范围限制在，的范围限制在，同时设置最小RTT探测周期的下限为7秒。

奖励函数的设计目标为最大化吞吐量，同时降低往返时延和丢包率，保证多流环境下带宽分配的公平性，强化学习模型的奖励函数设计为：

本发明实施例中，吞吐奖励权重系数取0.4，时延惩罚权重系数取0.2，拥塞惩罚权重系数取0.3，公平奖励权重系数取0.1。其中，是当前的吞吐量，即最近一个RTT交付的数据量；是Jain公平性指数。

对于吞吐奖励，通过将吞吐量归一化到占用瓶颈带宽的比例，激励算法向上收敛到链路的理论最大吞吐量；对于时延惩罚项，通过惩罚高时延，以减少队列堆积；对于拥塞惩罚项，结合ECN-Echo标记比例和丢包率，并指定较高权重，使得算法可以快速响应网络拥塞；对于公平奖励项，通过计算Jain公平性指数，激励算法在多流环境下，公平分配链路带宽，避免单一数据流垄断链路带宽，保证多流环境下的公平性。

本发明的进一步改进在于：在所述步骤5中，结合离线学习和在线学习对模型进行训练和更新。离线学习阶段，使用ns3网络仿真器和ns3-gym强化学习工具包模拟网络环境，使用PPO近端策略优化算法训练强化学习模型。设置带宽波动、往返时延抖动、多流竞争等多种挑战；模拟数据中心网络、无线网络等多种场景，提高模型的泛化能力。在线学习阶段，将强化学习模型部署到用户态的独立进程中。强化学习进程可以与内核BBR协议栈通信，周期性地从内核BBR协议栈中获取实时链路状态特征，包括链路瓶颈带宽和最小RTT、当前的交付速率、近期的RTT抖动和丢包率等，以及ECN-Echo标记比例，一并输入到强化学习模型。强化学习模型输出动作后，传递给内核BBR协议栈。强化学习进程使用滑动窗口缓存近期样本，定期触发小批量策略更新，快速响应网络环境变化。

为了保证在线学习的安全性，本发明引入动作幅度限制、异常策略检测和自动回滚机制。具体的，通过约束输出动作的变化范围，避免策略更新导致BBR关键参数突变。同时对比策略更新前后的网络传输性能，如果发现网络传输性能低于预期，则自动回滚到稳定版本策略，避免异常的策略更新导致网络传输性能明显下降。

本发明的有益效果是：

本发明对比传统的BBR拥塞控制协议，结合强化学习和ECN显式拥塞通知机制，实现了对BBR关键参数的动态优化。通过引入ECN显式拥塞通知机制，允许接收端实时反馈拥塞信息，发送端使用拥塞信息对拥塞窗口增益、最小RTT探测周期进行细粒度调整。

本发明并非单一的使用接收端反馈的拥塞信息减小拥塞窗口，而是同时将BBR实时探测的链路状态特征和接收端反馈的拥塞信息，一并输入到强化学习模型，进一步增强BBR对链路状态的实时感知能力和快速拥塞响应能力。

本发明的奖励函数综合考虑了吞吐量、往返时延和丢包率，可以适当调整奖励项或惩罚项的权重，以权衡不同的网络传输性能指标。通过结合离线学习和在线学习，快速响应网络环境变化。引入动作幅度限制、异常策略检测和自动回滚机制，保证在线学习的安全性，避免异常的策略更新导致网络传输性能明显下降。

本发明适用于数据中心网络、无线网络等多种场景，纯软件实现，不依赖特定的硬件，易于在网络协议栈中部署。

**附图说明**

图1是本发明结合强化学习和ECN显式拥塞通知机制，动态优化BBR关键参数的流程图。

图2是本发明结合强化学习和ECN显式拥塞通知机制，动态优化BBR关键参数的结构示意图。

**具体实施方式**

以下将以图式揭露本发明的实施方式，为明确说明起见，许多实务上的细节将在以下叙述中一并说明。然而，应了解到，这些实务上的细节不应用以限制本发明。也就是说，在本发明的部分实施方式中，这些实务上的细节是非必要的。

本发明是一种对BBR关键参数的动态优化方法，该方法通过基于强化学习和ECN显式拥塞通知机制的响应式模型实现，该模型包括：

BBR状态机模块：构建BBR状态机，作为后续强化学习模型的基础调度框架。

链路状态探测模块：BBR会持续探测链路状态特征，包括链路瓶颈带宽和最小RTT、当前的交付速率、近期的RTT抖动和丢包率等。特别的，发送端会持续维护指数加权移动平均的RTT抖动，用于响应链路状态的突发改变。

接收端CE标记处理模块：接收端收到携带CE标记的数据包后，通过在ACK中设置ECN-Echo标记，将拥塞信息反馈给发送端。

发送端ECN-Echo标记处理模块：发送端收到携带ECN-Echo标记的ACK后，计算每个RTT内收到的ACK中，携带ECN-Echo标记的ACK数量占ACK总数量的比例，用于对拥塞窗口增益和最小RTT探测周期进行细粒度调整。

强化学习策略模块：输入BBR实时采集的链路状态特征和ECN-Echo标记比例，计算奖励值，动态生成对BBR关键参数，包括拥塞窗口增益、最小RTT探测周期的调整策略。

奖励值计算模块：使用奖励函数周期性计算链路状态特征和ECN-Echo标记比例更新后的奖励值，奖励值计算模块包括吞吐奖励子模块、时延惩罚子模块、拥塞惩罚子模块和公平奖励子模块。

如图1所示，本发明是一种结合强化学习和ECN显式拥塞通知机制，对BBR关键参数的动态优化方法，该方法包括以下步骤：

步骤1、构建BBR状态机，作为后续强化学习模型的基础调度框架。BBR状态机包括四个阶段：Startup启动阶段、Drain队列排空阶段、ProbeBW带宽探测阶段、ProbeRTT最小RTT探测阶段。启动阶段指数增大发送速率，以探测链路的瓶颈带宽；队列排空阶段减小发送速率，以排空启动阶段形成的队列积压；带宽探测阶段是稳态阶段，持续探测链路的瓶颈带宽，并周期性进入最小RTT探测阶段，使用极低的发送速率排空瓶颈缓冲区的积压, 更新最小RTT。

步骤2、构建发送端链路状态探测模块。发送端实时探测链路瓶颈带宽和最小RTT、当前的交付速率、近期的RTT抖动和丢包率等。特别的，发送端持续维护指数加权移动平均的RTT抖动，用于响应链路状态的突发改变，指数加权移动平均的RTT抖动的计算公式为：

其中，是指数加权移动平均的RTT抖动；是平滑因子；是当前RTT的测量值；是BBR持续维护的最小RTT。

步骤3、构建接收端CE标记处理模块。如图2所示，支持ECN的交换机会实时测量瞬时队列长度，当瞬时队列长度超过阈值时，交换机会在数据包的IP头部中设置CE标记，表示拥塞经历。接收端收到携带CE标记的数据包后，通过在ACK的TCP头部中设置ECN-Echo标记，将拥塞信息反馈给发送端。

步骤4、构建发送端ECN-Echo标记处理模块。如图2所示，发送端收到携带ECN-Echo标记的ACK后，计算每个RTT内收到的ACK中，携带ECN-Echo标记的ACK数量占ACK总数量的比例，该比例一方面对拥塞窗口增益、最小RTT探测周期进行细粒度调整，另一方面作为强化学习模型的输入之一。细粒度调整的公式为：

其中，是一个RTT内，发送端收到的ACK中，携带ECN-Echo标记的ACK数量，占ACK总数量的比例；是拥塞窗口增益；是最小RTT探测周期；和是BBR持续维护的最大交付速率，也即链路的瓶颈带宽和最小RTT；是拥塞窗口大小。

步骤5、在BBR状态机中嵌入强化学习模型，强化学习模型作用于BBR状态机的带宽探测阶段和最小RTT探测阶段，通过进程间通信与内核BBR协议栈交互。

步骤6、使用ns3和ns3-gym进行预训练，模拟网络环境。使用PPO近端策略优化算法训练强化学习模型。设置带宽波动、往返时延抖动、多流竞争等多种挑战；模拟数据中心网络、无线网络等多种场景，提高模型的泛化能力。

步骤7、将强化学习模型部署到用户态的独立进程。发送端和接收端运行BBR协议，发送端运行前可以加载对应网络环境下的预训练参数，也可以使用默认参数：拥塞窗口增益默认值是2，最小RTT探测周期默认值是10秒。强化学习进程周期性地从内核BBR协议栈中获取实时链路状态特征，包括链路瓶颈带宽和最小RTT、当前的交付速率、近期的RTT抖动和丢包率等，以及ECN-Echo标记比例，一并输入到强化学习模型。强化学习模型使用的奖励函数公式为：

其中，是当前的吞吐量，即最近一个RTT交付的数据量；是近期10个RTT内的平均丢包率；是Jain公平性指数。奖励函数的第一项是吞吐奖励子模块，奖励高吞吐，激励算法向上收敛到链路的理论最大吞吐量；第二项是时延惩罚子模块，惩罚高时延，以减少队列堆积；第三项是拥塞惩罚子模块，通过指定较高权重以快速响应网络拥塞；第四项是公平奖励子模块，奖励多流环境下公平分配带宽资源。

步骤8、结合强化学习和ECN机制动态优化BBR关键参数。强化学习模型输出动作后，传递给内核BBR协议栈。强化学习进程使用滑动窗口缓存近期样本，定期触发小批量策略更新，快速响应网络环境变化。通过约束输出动作的变化范围，避免策略更新导致BBR关键参数突变。同时对比策略更新前后的网络传输性能，如果发现网络传输性能低于预期，则自动回滚到稳定版本策略，避免异常的策略更新导致网络传输性能明显下降。

本发明结合强化学习和ECN显式拥塞通知机制，动态优化BBR协议的拥塞窗口增益和最小RTT探测周期，基于实时链路状态特征和ECN-Echo标记比例，设计奖励函数，结合离线学习和在线学习，实现在高动态的网络环境下对BBR关键参数的动态优化。适用于数据中心网络、无线网络等多种场景，纯软件实现，不依赖特定的硬件，易于在网络协议栈中部署。

以上所述仅为本发明的实施方式而已，并不用于限制本发明。对于本领域技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原理的内所作的任何修改、等同替换、改进等，均应包括在本发明的权利要求范围之内。

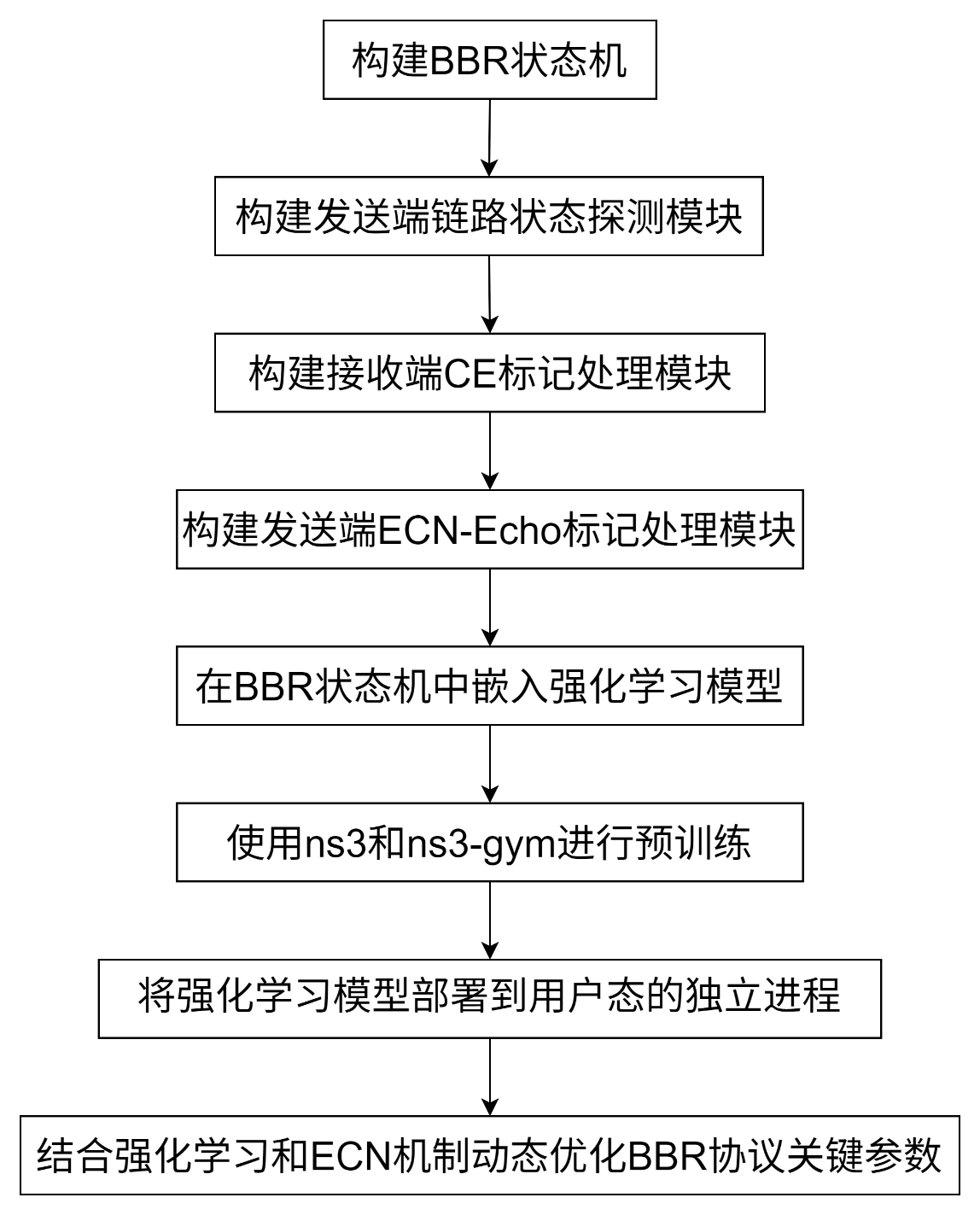


图1



图2