(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111683409 A
(43)申请公布日 2020.09.18

(21)申请号 202010506834.5

(22)申请日 2020.06.05

(71)申请人 上海特金无线技术有限公司
地址 201114 上海市闵行区新骏环路245号第6层E612室

(72)发明人 李瀚 姜维 姜化京

(74)专利代理机构 上海慧晗知识产权代理事务所(普通合伙) 31343
代理人 李茂林

(51)Int.Cl.
H04W 72/04(2009.01)
H04W 72/12(2009.01)
G05D 1/10(2006.01)

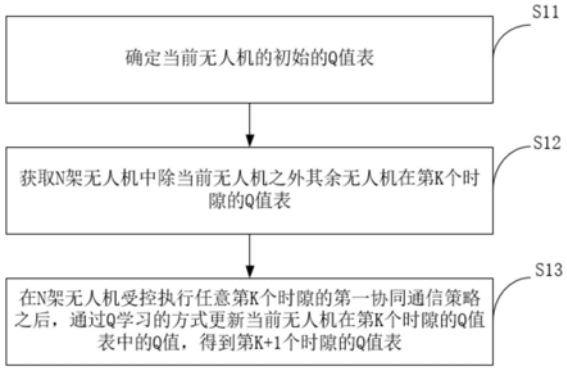
权利要求书3页 说明书11页 附图5页

(54)发明名称

多无人机协同通信Q值表的学习方法、调度方法及装置

(57)摘要

本发明提供了一种多无人机协同通信调度的Q值表的学习方法、多无人机协同通信调度方法、Q值表的学习装置、电子设备及存储介质。其中Q值表的学习方法包括:确定当前无人机的初始的Q值表;Q值表记载了对应的无人机在每一种环境状态下,N架无人机执行不同的协同通信策略对应的Q值;其中,不同的环境状态表征了目标频段的信号的不同受干扰情况;每个协同通信策略表征了N架无人机中各架无人机的通信策略的一种组合;获取N架无人机中除当前无人机之外其余无人机在第K个时隙的Q值表;通过Q学习的方式更新当前无人机在第K个时隙的Q值表中的Q值,得到第K+1个时隙的Q值表。本发明提供的Q值表的学习方法能够提高N架无人机整体的调度成功率。



CN 111683409 A

1.一种多无人机协同通信调度的Q值表的学习方法，应用于N架无人机中任意之一的当前无人机，其中的N为大于或等于2的整数，其特征在于，所述的学习方法，包括：

确定所述当前无人机的初始的Q值表；所述Q值表记载了对应的无人机在每一种环境状态下，所述N架无人机执行不同的协同通信策略对应的Q值；其中，不同的环境状态表征了通信频段的信号的不同受干扰情况；每个协同通信策略表征了所述N架无人机中各架无人机的通信策略的一种组合；

获取所述N架无人机中除所述当前无人机之外其余无人机在任意第K个时隙的Q值表；

在所述N架无人机受控执行所述第K个时隙的第一协同通信策略之后，通过Q学习的方式更新所述当前无人机在所述第K个时隙的Q值表中的Q值，得到第K+1个时隙的Q值表；所述第K+1个时隙的Q值表是根据所述第一协同通信策略、所述第K个时隙的第一环境状态、第K+1个时隙的第二环境状态，以及所述其余无人机在第K个时隙的Q值表更新的。

2.根据权利要求1所述的方法，其特征在于，通过Q学习的方式更新所述当前无人机在所述第K个时隙的Q值表中的Q值，包括：

确定所述当前无人机在第K个时隙对应的调度回报值；所述调度回报值表征了所述当前无人机执行所述第一协同通信策略中相应策略后是否成功通信；

根据第K个时隙的N张Q值表以及所述第二环境状态，在所有协同通信策略中确定第二协同通信策略；其中，所述第二协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的；所述第二协同通信策略的Q值统计值是针对于所述第K个时隙的N张Q值表中所述第二协同通信策略与所述第二环境状态所对应的N个Q值进行统计而得到的；

更新所述当前无人机在第K个时隙的Q值表中的Q值，得到所述当前无人机的第K+1个时隙对应的Q值表。

3.根据权利要求2所述的方法，其特征在于，所述当前无人机的第K个时隙对应的Q值表中更新的Q值表征为：

$$Q_n^K(s_K,a)=Q_n^K(s_K,a)+\lambda\left[r_n+\gamma Q_n^K\left(s_{K+1},a^*\right)-Q_n^K(s_K,a)\right]$$；

其中，n为所述当前无人机的编号，Q_n^K为所述当前无人机在第K个时隙的Q值表，$\lambda\in(0,1)$为学习率；$\gamma\in(0,1)$为算法折扣因子；r_n为所述调度回报值；a为所述第一协同通信策略，s_K为所述第一环境状态，s_{K+1}为所述第二环境状态，a^*为所述第二协同通信策略。

4.根据权利要求2所述的方法，其特征在于，所述第二协同通信策略的Q值统计值为所述第K个时隙的N张Q值表中所述第二协同通信策略与所述第二环境状态所对应的N个Q值之和。

5.根据权利要求2-4任一项所述的方法，其特征在于，在所述N架无人机受控执行任意第K个时隙的第一协同通信策略之前，还包括：

以概率ε在两个策略确定方式中选择第一策略确定方式后，利用所述第一策略确定方式确定所述第一协同通信策略；或者：

以概率1-ε，在两个策略确定方式中选择第二策略确定方式后，利用所述第二策略确定方式确定所述第一协同通信策略；

利用第一策略确定方式确定所述第一协同通信策略包括：

随机确定N架无人机的通信策略的一种组合作为所述第一协同通信策略；

利用第二策略确定方式确定所述第一协同通信策略，包括：

根据第K-1个时隙的N张Q值表以及所述第一环境状态，在所有协同通信策略中确定所述第一协同通信策略；所述第一协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的；所述第一协同通信策略的Q值统计值是针对于所述第K-1个时隙的N张Q值表中所述第一协同通信策略与所述第一环境状态所对应的N个Q值进行统计而得到的。

6.一种多无人机协同通信调度方法，应用于N架待调度无人机中任意之一的当前待调度无人机，其特征在于，所述待调度无人机中存储有权利要求1至5任一项所确定的Q值表，包括：

确定所述N架待调度无人机在当前时隙的当前环境状态；

根据所述当前环境状态以及权利要求1-5任意一项所述的多无人机协同通信调度的Q值表的学习方法确定的Q值表，确定当前协同通信策略；

执行所述当前协同通信策略中所述当前待调度无人机所对应的策略，以在目标频段通信或保持静默。

7.一种多无人机协同通信调度的Q值表的学习装置，应用于N架无人机中任意之一的当前无人机，其中的N为大于或等于2的整数，其特征在于，所述的学习装置，包括：

初始Q值表确定模块，用于确定所述当前无人机的初始的Q值表；所述Q值表记载了对应的无人机在每一种环境状态下，所述N架无人机执行不同的协同通信策略对应的Q值；其中，不同的环境状态表征了目标频段的信号的不同受干扰情况；每个协同通信策略表征了所述N架无人机中各架无人机的通信策略的一种组合；

Q值表获取模块，用于在所述N架无人机受控执行任意第K个时隙的第一协同通信策略之后，获取所述N架无人机中除所述当前无人机之外其余无人机在所述第K个时隙的Q值表；

更新模块，用于通过Q学习的方式更新所述当前无人机在所述第K个时隙的Q值表中的Q值，得到第K+1个时隙的Q值表；所述第K+1个时隙的Q值表是根据所述第一协同通信策略、所述第K个时隙的第一环境状态、第K+1个时隙的第二环境状态，以及所述其余无人机在第K个时隙的Q值表更新的。

8.根据权利要求7所述的装置，其特征在于，所述更新模块，包括：

反馈单元，用于确定所述当前无人机在第K个时隙对应的调度回报值；所述调度回报值表征了所述当前无人机执行所述第一协同通信策略中相应策略后是否成功通信；

第二协同通信策略确定单元，用于根据第K个时隙的N张Q值表以及所述第二环境状态，在所有协同通信策略中确定第二协同通信策略；其中，所述第二协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的；所述第二协同通信策略的Q值统计值是针对于所述第K个时隙的N张Q值表中所述第二协同通信策略与所述第二环境状态所对应的N个Q值进行统计而得到的；

第一计算单元，用于更新所述当前无人机在第K个时隙的Q值表中的Q值，得到所述当前无人机的第K+1个时隙对应的Q值表。

9.根据权利要求8所述的装置，其特征在于，所述当前无人机的第K个时隙对应的Q值表中更新的Q值表征为：

$$Q_n^K(s_K,a)=Q_n^K(s_K,a)+\lambda\left[r_n+\gamma Q_n^K\left(s_{K+1},a^*\right)-Q_n^K(s_K,a)\right]$$;

其中，n为所述当前无人机的编号，Q_n^K为所述当前无人机在第K个时隙的Q值表，λ∈(0,1)为学习率；γ∈(0,1)为算法折扣因子；r_n为所述调度回报值；a为所述第一协同通信策略，s_K为所述第一环境状态，s_{K+1}为所述第二环境状态，a^*为所述第二协同通信策略。

10.根据权利要求8所述的装置，其特征在于，所述第二协同通信策略的Q值统计值为所述第K个时隙的N张Q值表中所述第二协同通信策略与所述第二环境状态所对应的N个Q值之和。

11.根据权利要求8-10任一项所述的装置，其特征在于，还包括第一协同通信策略确定模块，用于：

以概率ε在两个策略确定方式中选择第一策略确定方式后，利用所述第一策略确定方式确定所述第一协同通信策略；或者：

以概率1-ε，在两个策略确定方式中选择第二策略确定方式后，利用所述第二策略确定方式确定所述第一协同通信策略；

利用第一策略确定方式确定所述第一协同通信策略包括：

随机确定N架无人机的通信策略的一种组合作为所述第一协同通信策略；

利用第二策略确定方式确定所述第一协同通信策略，包括：

根据第K-1个时隙的N张Q值表以及所述第一环境状态，在所有协同通信策略中确定所述第一协同通信策略；所述第一协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的；所述第一协同通信策略的Q值统计值是针对于所述第K-1个时隙的N张Q值表中所述第一协同通信策略与所述第一环境状态所对应的N个Q值进行统计而得到的。

12.一种电子设备，其特征在于，包括处理器与存储器，

所述存储器，用于存储代码和相关数据；

所述处理器，用于执行所述存储器中的代码用以实现权利要求1-5任一项所述的方法。

13.一种存储介质，其上存储有计算机程序，该程序被处理器执行时实现权利要求1-5任一项所述的方法。

多无人机协同通信Q值表的学习方法、调度方法及装置

技术领域

[0001] 本发明涉及无人机领域，并且更具体地，涉及一种多无人机协同通信调度的Q值表的学习方法、多无人机协同通信调度方法、Q值表的学习装置、电子设备及存储介质。

背景技术

[0002] 近年来，小型、低成本的无人机集群受到广泛关注，和单个无人机相比，多无人机集群在灵活性、容错性、协作性、任务多样性等方面都有明显优势，无人机集群的应用越来越广泛。例如，无人机集群中的各无人机可以充当接入节点，作为地面通信站点或通信设备的信号中继器。

[0003] 现有技术中，无人机集群协同执行的任务由控制中心统一指挥调度，控制中心生成控制指令发送给无人机集群，无人机集群中各架无人机根据控制指令执行相应的动作。

[0004] 可见，现有技术中，无人机集群中各架无人机的动作仅由控制中心提前设定，其中的每架无人机不能根据其自身飞行环境以及其余无人机的状态独立作出决策，从而导致无人机集群调度成功率低。

发明内容

[0005] 本发明提供了一种多无人机协同通信调度的Q值表的学习方法、多无人机协同通信调度方法、Q值表的学习装置、电子设备及存储介质，以解决现有的多无人机调度技术中每架无人机不能根据其自身飞行环境以及其余无人机的状态独立作出决策，多无人机的整体调度成功率低的问题。

[0006] 根据本发明的第一方面，提供了一种多无人机协同通信调度的Q值表的学习方法，应用于N架无人机中任意之一的当前无人机，其中的N为大于或等于2的整数，所述的学习方法，包括：

[0007] 确定所述当前无人机的初始的Q值表；所述Q值表记载了对应的无人机在每一种环境状态下，所述N架无人机执行不同的协同通信策略对应的Q值；其中，不同的环境状态表征了目标频段的信号的不同受干扰情况；每个协同通信策略表征了所述N架无人机中各架无人机的通信策略的一种组合；

[0008] 在所述N架无人机受控执行任意第K个时隙的第一协同通信策略之后，获取所述N架无人机中除所述当前无人机之外其余无人机在所述第K个时隙的Q值表；

[0009] 通过Q学习的方式更新所述当前无人机在所述第K个时隙的Q值表中的Q值，得到第K+1个时隙的Q值表；所述第K+1个时隙的Q值表是根据所述第一协同通信策略、所述第K个时隙的第一环境状态、第K+1个时隙的第二环境状态，以及所述其余无人机在第K个时隙的Q值表更新的。

[0010] 可选的，通过Q学习的方式更新所述当前无人机在所述第K个时隙的Q值表中的Q值，包括：

[0011] 确定所述当前无人机在第K个时隙对应的调度回报值；所述调度回报值表征了所

述当前无人机执行所述第一协同通信策略中相应策略后是否成功通信；

[0012] 根据第K个时隙的N张Q值表以及所述第二环境状态，在所有协同通信策略中确定第二协同通信策略；其中，所述第二协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的；所述第二协同通信策略的Q值统计值是针对于所述第K个时隙的N张Q值表中所述第二协同通信策略与所述第二环境状态所对应的N个Q值进行统计而得到的；

[0013] 更新所述当前无人机在第K个时隙的Q值表中的Q值，得到所述当前无人机的第K+1个时隙对应的Q值表。

[0014] 可选的，所述当前无人机的第K个时隙对应的Q值表中更新的Q值表征为：

[0015] $$Q_n^K(s_K,a)=Q_n^K(s_K,a)+\lambda\left[r_n+\gamma Q_n^K\left(s_{K+1},a^*\right)-Q_n^K(s_K,a)\right]$$；

[0016] 其中，n为所述当前无人机的编号，Q_n^K为所述当前无人机在第K个时隙的Q值表，$\lambda\in(0,1)$为学习率；$\gamma\in(0,1)$为算法折扣因子；r_n为所述调度回报值；a为所述第一协同通信策略，s_K为所述第一环境状态，s_{K+1}为所述第二环境状态，a^*为所述第二协同通信策略。

[0017] 可选的，所述第二协同通信策略的Q值统计值为所述第K个时隙的N张Q值表中所述第二协同通信策略与所述第二环境状态所对应的N个Q值之和。

[0018] 可选的，在所述N架无人机受控执行任意第K个时隙的第一协同通信策略之前，还包括：

[0019] 以概率ε在两个策略确定方式中选择第一策略确定方式后，利用所述第一策略确定方式确定所述第一协同通信策略；或者：

[0020] 以概率1-ε，在两个策略确定方式中选择第二策略确定方式后，利用所述第二策略确定方式确定所述第一协同通信策略；

[0021] 利用第一策略确定方式确定所述第一协同通信策略包括：

[0022] 随机确定N架无人机的通信策略的一种组合作为所述第一协同通信策略；

[0023] 利用第二策略确定方式确定所述第一协同通信策略，包括：

[0024] 根据第K-1个时隙的N张Q值表以及所述第一环境状态，在所有协同通信策略中确定所述第一协同通信策略；所述第一协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的；所述第一协同通信策略的Q值统计值是针对于所述第K-1个时隙的N张Q值表中所述第一协同通信策略与所述第一环境状态所对应的N个Q值进行统计而得到的。

[0025] 根据本发明的第二方面，提供了一种多无人机协同通信调度方法，应用于N架待调度无人机中任意之一的当前待调度无人机，所述待调度无人机中存储有本发明第一方面及其可选方案涉及的方法所确定的Q值表，包括：

[0026] 确定所述N架待调度无人机在当前时隙的当前环境状态；

[0027] 根据所述当前环境状态以及本发明第一方面及其可选方案涉及的方法所确定的Q值表，确定当前协同通信策略；

[0028] 执行所述当前协同通信策略中所述当前待调度无人机所对应的策略，以在目标频段通信或保持静默。

[0029] 根据本发明的第三方面，提供了一种多无人机协同通信调度的Q值表的学习装置，应用于N架无人机中任意之一的当前无人机，其中的N为大于或等于2的整数，所述的学习装置，包括：

[0030] 初始Q值表确定模块，用于确定所述当前无人机的初始的Q值表；所述Q值表记载了对应的无人机在每一种环境状态下，所述N架无人机执行不同的协同通信策略对应的Q值；其中，不同的环境状态表征了目标频段的信号的不同受干扰情况；每个协同通信策略表征了所述N架无人机中各架无人机的通信策略的一种组合；

[0031] Q值表获取模块，用于在所述N架无人机受控执行任意第K个时隙的第一协同通信策略之后，获取所述N架无人机中除所述当前无人机之外其余无人机在所述第K个时隙的Q值表；

[0032] 更新模块，用于通过Q学习的方式更新所述当前无人机在所述第K个时隙的Q值表中的Q值，得到第K+1个时隙的Q值表；所述第K+1个时隙的Q值表是根据所述第一协同通信策略、所述第K个时隙的第一环境状态、第K+1个时隙的第二环境状态，以及所述其余无人机在第K个时隙的Q值表更新的。

[0033] 可选的，所述更新模块，包括：

[0034] 反馈单元，用于确定所述当前无人机在第K个时隙对应的调度回报值；所述调度回报值表征了所述当前无人机执行所述第一协同通信策略中相应策略后是否成功通信；

[0035] 第二协同通信策略确定单元，用于根据第K个时隙的N张Q值表以及所述第二环境状态，在所有协同通信策略中确定第二协同通信策略；其中，所述第二协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的；所述第二协同通信策略的Q值统计值是针对于所述第K个时隙的N张Q值表中所述第二协同通信策略与所述第二环境状态所对应的N个Q值进行统计而得到的；

[0036] 第一计算单元，用于更新所述当前无人机在第K个时隙的Q值表中的Q值，得到所述当前无人机的第K+1个时隙对应的Q值表。

[0037] 可选的，所述当前无人机的第K个时隙对应的Q值表中更新的Q值表征为：

[0038] $$Q_n^K(s_K,a)=Q_n^K(s_K,a)+\lambda\left[r_n+\gamma Q_n^K(s_{K+1},a^*)-Q_n^K(s_K,a)\right]$$；

[0039] 其中，n为所述当前无人机的编号，Q_n^K为所述当前无人机在第K个时隙的Q值表，$\lambda\in(0,1)$为学习率；$\gamma\in(0,1)$为算法折扣因子；r_n为所述调度回报值；a为所述第一协同通信策略，s_K为所述第一环境状态，s_{K+1}为所述第二环境状态，a^*为所述第二协同通信策略。

[0040] 可选的，所述第二协同通信策略的Q值统计值为所述第K个时隙的N张Q值表中所述第二协同通信策略与所述第二环境状态所对应的N个Q值之和。

[0041] 可选的，所述的装置，还包括第一协同通信策略确定模块，用于：

[0042] 以概率ε在两个策略确定方式中选择第一策略确定方式后，利用所述第一策略确定方式确定所述第一协同通信策略；或者：

[0043] 以概率1-ε，在两个策略确定方式中选择第二策略确定方式后，利用所述第二策略确定方式确定所述第一协同通信策略；

[0044] 利用第一策略确定方式确定所述第一协同通信策略包括：

[0045] 随机确定N架无人机的通信策略的一种组合作为所述第一协同通信策略；

[0046] 利用第二策略确定方式确定所述第一协同通信策略，包括：

[0047] 根据第K-1个时隙的N张Q值表以及所述第一环境状态，在所有协同通信策略中确定所述第一协同通信策略；所述第一协同通信策略的Q值统计值是对应时隙所有协同通信

策略中最大的;所述第一协同通信策略的Q值统计值是针对于所述第K-1个时隙的N张Q值表中所述第一协同通信策略与所述第一环境状态所对应的N个Q值进行统计而得到的。

[0048] 根据本发明的第四方面,提供了一种电子设备,包括处理器与存储器,

[0049] 所述存储器,用于存储代码和相关数据;

[0050] 所述处理器,用于执行所述存储器中的代码用以实现本发明第一方面及其可选方案涉及的方法。

[0051] 根据本发明的第五方面,提供了一种存储介质,其上存储有计算机程序,该程序被处理器执行时实现本发明第一方面及其可选方案涉及的方法。

[0052] 本发明提供的一种多无人机协同通信调度的Q值表的学习方法、多无人机协同通信调度方法、Q值表的学习装置、电子设备及存储介质,其中通过学习确定的当前无人机的Q值表记载了对应的无人机在每一种环境状态下,N架无人机执行不同的协同通信策略对应的Q值,相对于现有技术,当前无人机可根据对应的已通过学习确定的Q值表独立作出决策;由于不同的环境状态表征了目标频段的信号的不同受干扰情况,每个协同通信策略表征了N架无人机中各架无人机的通信策略的一种组合,进而,当前无人机根据已确定的Q值表作出决策时,已经考虑了当前无人机的飞行环境(目标频段的信号受干扰情况)以及其余无人机的状态(其余无人机的通信策略)。相对于现有技术,N架无人机中任意之一的当前无人机确定通信策略时均增加了作决策时考虑的因素,故而,本发明提供的一种多无人机协同通信调度的Q值表的学习方法、多无人机协同通信调度方法、Q值表的学习装置、电子设备及存储介质,能够提高N架无人机整体的调度成功率。

附图说明

[0053] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0054] 图1是本发明一实施例中多无人机协同通信调度的Q值表的学习场景图;

[0055] 图2是本发明一实施例中多无人机协同通信调度的Q值表的学习方法的流程图一;

[0056] 图3是本发明一实施例中多无人机协同通信调度的Q值表的学习方法的流程图二;

[0057] 图4是本发明一实施例中调度成功率随迭代次数变化仿真图;

[0058] 图5是本发明一实施例中多无人机协同通信调度方法的流程图;

[0059] 图6是本发明一实施例中多无人机协同通信调度的Q值表的学习装置的模块示意图一;

[0060] 图7是本发明一实施例中多无人机协同通信调度的Q值表的学习装置的模块示意图二;

[0061] 图8是本发明一实施例中多无人机协同通信调度的Q值表的学习装置的模块示意图三;

[0062] 图9是本发明一实施例中电子设备的构造示意图。

[0063] 附图标记说明:

[0064] 11-N架无人机;

[0065] 1101-当前无人机；

[0066] 12-干扰设备；

[0067] 13-地面通信设备；

[0068] 14-目标通信设备；

[0069] 21-初始Q值表确定模块；

[0070] 22-Q值表获取模块；

[0071] 23-更新模块；

[0072] 2301-反馈单元；

[0073] 2302-第二协同通信策略确定单元；

[0074] 2303-第一计算单元；

[0075] 24-第一协同通信策略确定模块；

[0076] 31-处理器；

[0077] 32-总线；

[0078] 33-存储器。

具体实施方式

[0079] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0080] 本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”“第四”等(如果存在)是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本发明的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0081] 下面以具体地实施例对本发明的技术方案进行详细说明。下面这几个具体的实施例可以相互结合，对于相同或相似的概念或过程可能在某些实施例不再赘述。

[0082] 图1是本发明一实施例中多无人机协同通信调度的Q值表的学习场景图。

[0083] 请参考图1，该场景所关联的设备可包括:N架无人机11、当前无人机1101、干扰设备12、地面通信设备13，以及目标通信设备14。

[0084] 图1所示场景中，N架无人机11充当接入节点，地面通信设备13(例如，通信基站、临时通信站点、移动通信终端、无线发射器等)通过无人机作为信号中继站与目标通信设备14建立通信连接，当然，地面通信设备13的目标通信设备14也可以是无人机本身。在无人机与地面通信设备13建立连接前，无人机需要向地面通信设备13发送信标信号，以向地面通信设备13声明无人机的存在并为用户接入无人机提供通道，当地面通信设备13正确解调出信标信号之后，才能够接入相应的无人机。

[0085] N架无人机11执行调度任务时，其中的各架无人机按照同步且相同的时隙完成调

度动作，若N架无人机11中其中一架无人机(例如，当前无人机1101)与地面通信设备13建立了通信连接，则认为N架无人机11在对应时隙的调度任务成功，否则，认为调度失败。由于频谱环境存在干扰设备12发送的信号(以下称为干扰信号)，干扰信号的频段如果与无人机的信号(例如信标信号)的频段相同或有重叠，地面通信设备13会收到发生冲突的信号导致地面通信设备13无法正确解调信号，进而，地面通信设备13无法与其中任何一架无人机建立通信连接；N架无人机11中还可能存在至少两架无人机同时发送信标信号的情况，同样的，这也会导致地面通信设备13无法正确解调信号。

[0086] 图2是本发明一实施例中多无人机协同通信调度的Q值表的学习方法的流程图一。

[0087] 请参考图1以及图2，一种实施方式中，多无人机协同通信调度的Q值表的学习方法，应用于N架无人机11中任意之一的当前无人机1101，其中的N为大于或等于2的整数，该学习方法，包括：

[0088] S11：确定当前无人机1101的初始的Q值表；Q值表记载了对应的无人机在每一种环境状态下，N架无人机11执行不同的协同通信策略对应的Q值；其中，不同的环境状态表征了目标频段的信号的不同受干扰情况；每个协同通信策略表征了N架无人机11中各架无人机的通信策略的一种组合；

[0089] S12：获取N架无人机11中除当前无人机1101之外其余无人机在第K个时隙的Q值表；

[0090] S13：在N架无人机11受控执行任意第K个时隙的第一协同通信策略之后，通过Q学习的方式更新当前无人机1101在第K个时隙的Q值表中的Q值，得到第K+1个时隙的Q值表；第K+1个时隙的Q值表是根据第一协同通信策略、第K个时隙的第一环境状态、第K+1个时隙的第二环境状态，以及其余无人机在第K个时隙的Q值表更新的。

[0091] 以上方案中，通过学习确定的当前无人机1101的Q值表记载了对应的无人机在每一种环境状态下，N架无人机11执行不同的协同通信策略对应的Q值，相对于现有技术，当前无人机1101可根据对应的已通过学习确定的Q值表独立作出决策；由于不同的环境状态表征了目标频段的信号的不同受干扰情况，每个协同通信策略表征了N架无人机11中各架无人机的通信策略的一种组合，进而，当前无人机1101根据已确定的Q值表作出决策时，已经考虑了当前无人机1101的飞行环境(目标频段的信号受干扰情况)以及其余无人机的状态(其余无人机的通信策略)。相对于现有技术，N架无人机11中任意之一的当前无人机1101确定通信策略时均增加了作决策时考虑的因素，故而，以上方案涉及的多无人机协同通信调度的Q值表的学习方法，能够提高N架无人机11整体的调度成功率。

[0092] 以上方案中，由于提高了N架无人机11整体的调度成功率，进而，减少了调度失败造成的无人机能耗损耗。

[0093] 本发明实施例中，目标频段可以理解为无人机与地面通信设备13之间的通信频段，进一步可以理解为，无人机发送的信标信号的频段。

[0094] 本发明实施例中，Q值表的学习方法应用于N架无人机11中任意之一的当前无人机1101可以理解为，N架无人机11中的各架无人机按照同步且相同的时隙采用相同的Q值表的学习方法确定与之相对应的Q值表，进而，在学习完成后，实际应用每张Q值表的无人机可以独立作出决策。

[0095] 本发明实施例中，环境状态可表征为：

[0096] $s_K=[o_K,o_{K-1},\cdots,o_{K-M+1}]$；其中，$s_K$为第K个时隙的环境状态，其可利用多个环境状态值的集合来表征；M为预设的历史时隙长度；o_K为环境状态值，表征第K个时隙是否存在目标频段的干扰信号。

[0097] 一种实施方式中，N架无人机11对目标频段进行感知，可以通过能量检测的方法判断是否存在干扰设备12在目标频段发送信号，获取当前频谱环境中的环境状态值。其他实施方式中，还可以通过匹配滤波或特性检测等方法获取当前频谱环境中的环境状态值。当检测到当前环境中存在干扰信号，则可以确定环境状态值o_K为1；当检测到当前环境中不存在干扰信号，则可以确定环境状态值o_K为0。

[0098] 本发明实施例中，协同通信策略可表征为：

[0099] $a_K^n=[c_1,c_2,\cdots,c_n,\cdots,c_N]$；其中，1≤n≤N，N为无人机的数量；$a_K^n$表征当前无人机1101对应的Q值表在环境状态$s_K$下对应的协同通信策略；$c_n$表征当前无人机1101在第K个时隙所需执行的通信策略。

[0100] 本发明实施例中，通信策略为向地面通信设备13发送信标信号以调度地面通信设备13，或：保持静默以避免与干扰信号或其他无人机发送的信标信号发生碰撞。

[0101] 针对于步骤S12，在实际应用场景中，N架无人机11中各架无人机可通过无人机的广播系统广播自己的Q值表，当前无人机1101接收其余N-1架无人机在第K个时隙广播的第K个时隙的Q值表，获取的其余N-1架无人机的Q值表用于步骤S13中更新当前无人机1101在第K个时隙的Q值表中的Q值。当前无人机1101可以在N架无人机11受控执行任意第K个时隙的第一协同通信策略之前，获取其余N-1架无人机的Q值表，也可以在第一协同通信策略执行之后获取。

[0102] 本发明实施例中，初始的Q值表可以是全0矩阵，大小为X×Y，具体可表征为：

[0103] $$Q_0=\begin{bmatrix}0 & \cdots & \cdots & 0\\ \vdots & \ddots & & \vdots\\ \vdots & & \ddots & \vdots\\ 0 & \cdots & \cdots & 0\end{bmatrix};$$

[0104] 其中，$X=2^M$；$Y=2^K$。

[0105] 图3是本发明一实施例中多无人机协同通信调度的Q值表的学习方法的流程图二。

[0106] 请参考图3，通过Q学习的方式更新当前无人机在第K个时隙的Q值表中的Q值，即步骤S13，包括：

[0107] S131：确定当前无人机在第K个时隙对应的调度回报值；调度回报值表征了当前无人机执行第一协同通信策略中相应策略后是否成功通信；

[0108] S132：根据第K个时隙的N张Q值表以及第二环境状态，在所有协同通信策略中确定第二协同通信策略；其中，第二协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的；第二协同通信策略的Q值统计值是针对于第K个时隙的N张Q值表中第二协同通信策略与第二环境状态所对应的N个Q值进行统计而得到的；

[0109] S133：更新当前无人机在第K个时隙的Q值表中的Q值，得到当前无人机的第K+1个

时隙对应的Q值表。

[0110] 一种实施方式中，所述当前无人机的第K个时隙对应的Q值表中更新的Q值表征为：

[0111] $$Q_n^K(s_K,a)=Q_n^K(s_K,a)+\lambda\left[r_n+\gamma Q_n^K(s_{K+1},a^*)-Q_n^K(s_K,a)\right]$$；

[0112] 其中，n为当前无人机的编号，Q_n^K为当前无人机在第K个时隙的Q值表，$\lambda\in(0,1)$为学习率；$\gamma\in(0,1)$为算法折扣因子；r_n为调度回报值；a为第一协同通信策略，s_K为第一环境状态，s_{K+1}为第二环境状态，a^*为第二协同通信策略。

[0113] 针对于其中的调度回报值，若当前无人机执行第一协同通信策略中相应策略后成功通信，则认为调度成功，此时调度回报值可设置为1；若当前无人机执行第一协同通信策略中相应策略没有成功通信，则认为调度失败，此时调度回报值可设置为0。需要说明的是，调度回报值表征了当前无人机执行第一协同通信策略中相应策略后是否成功通信，进一步可以理解为，调度回报值表征对当前无人机执行策略的一种反馈，调度回报值设置为1表征对调度成功给予的正反馈，调度回报值设置为0表征对调度成功给予的负反馈，其中正反馈与负反馈是相对来讲的，故而，其他实施例中，调度回报值可以不限于设置为1或者0，只要能表征出对调度成功与调度失败不同的反馈即可。

[0114] 本发明实施例中，所有协同通信策略可以理解为N架无人机11中各架无人机的通信策略所有可能的组合。例如，若各架无人机的通信策略包括发送信标信号或保持静默，则N架无人机11的所有协同通信策略有2^N种。

[0115] 以上方案中，通过查找每个协同通信策略在第二环境状态下，每张Q值表中对应的Q值，N张Q值表可得到N个Q值，之后对N个Q值进行统计分析，得到Q值统计值，选择最大的Q值统计值对应的协同通信策略为第二协同通信策略，最大的Q值统计值表征第二协同通信策略是在第K个时隙N架无人机11在第二环境状态下整体最优的选择，通过第二协同通信策略以及第二环境状态可以查询当前无人机的Q值表中对应的Q值作为当前无人机未来的奖励。

[0116] 一种实施方式中，第二协同通信策略的Q值统计值为第K个时隙的N张Q值表中第二协同通信策略与第二环境状态所对应的N个Q值之和。

[0117] 第二协同通信策略的确定方式可以表征为：

[0118] $$a^*=\arg\max_{a'}\sum_{n=1}^{N}Q_n(S_{K+1},a')$$；

[0119] 其中a′为所有协同通信策略中的任一协同通信策略。

[0120] 当然，在其他实施方式中，本领域技术人员可以根据N架无人机11中各架无人机在调度中的重要程度或者稳定性等因素针对每个Q值设置对应的权重值，采用加权统计的方式计算Q值的统计值。

[0121] 一种实施方式中，在N架无人机11受控执行任意第K个时隙的第一协同通信策略之前，还包括：

[0122] 以概率ε在两个策略确定方式中选择第一策略确定方式后，利用第一策略确定方式确定第一协同通信策略；或者：以概率1-ε，在两个策略确定方式中选择第二策略确定方式后，利用第二策略确定方式确定第一协同通信策略；

[0123] 利用第一策略确定方式确定第一协同通信策略包括：

[0124] 随机确定N架无人机11的通信策略的一种组合作为第一协同通信策略；

[0125] 利用第二策略确定方式确定第一协同通信策略，包括：

[0126] 根据第K-1个时隙的N张Q值表以及第一环境状态，在所有协同通信策略中确定第一协同通信策略；第一协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的；第一协同通信策略的Q值统计值是针对于第K-1个时隙的N张Q值表中第一协同通信策略与第一环境状态所对应的N个Q值进行统计而得到的。

[0127] 本发明实施例中，其中$\varepsilon \in (0.01,1)$。

[0128] 本发明实施例中，随机确定N架无人机11的通信策略的一种组合作为第一协同通信策略可以理解为，N架无人机11中各架无人机随机选择发送信标信号或保持静默作为通信策略，各架无人机随机选择的通信策略的组合在一起作为第一协同通信策略。

[0129] 本发明实施例中，根据第K-1个时隙的N张Q值表以及第一环境状态，在所有协同通信策略中确定第一协同通信策略的方式与本发明实施例步骤S132中确定第二协同通信策略的方式及其可选方案基本相同，此处不再赘述。

[0130] 在其他实施方式中，还可以通过玻尔兹曼随机策略或者高斯策略确定第一协同通信策略。

[0131] 图4是本发明一实施例中调度成功率随迭代次数变化仿真图。

[0132] 假设无人机的数量N＝3，历史时隙长度M＝5，每架无人机可能的通信策略包括发送信标或保持静默，环境状态值为0或1，则Q值表的大小为32×8。图4是本发明一实施例中调度成功率随迭代次数变化仿真图，其横坐标表征迭代次数，纵坐标表征成功率，从图4可以看出，在算法初期，N架无人机11整体的调度成功率只有24％，随着算法的迭代，无人机渐渐学习出优选的策略，调度成功率趋于收敛，最终可达到70％左右，证实了算法的有效性。

[0133] 图5是本发明一实施例中多无人机协同通信调度方法的流程图。

[0134] 请参考图5，一种多无人机协同通信调度方法，应用于N架待调度无人机中任意之一的当前待调度无人机，待调度无人机中存储有本发明第一方面及其可选方案涉及的方法所确定的Q值表，包括：

[0135] S21：确定N架待调度无人机在当前时隙的当前环境状态；

[0136] S22：根据当前环境状态以及本发明第一方面及其可选方案涉及的方法所确定的Q值表，确定当前协同通信策略；

[0137] S23：执行当前协同通信策略中当前待调度无人机所对应的策略，以在目标频段通信或保持静默。

[0138] 以上方案中，由于N架待调度无人机均采用本发明第一方面及其可选方案涉及的方法确定的对应的Q值表，使得每架待调度无人机在根据当前环境状态以及对应的Q值表独立作决策时，均考虑了飞行时频谱环境的变化以及其余N-1架待调度无人机的决策，有效提高了对地面通信设备的调度成功率，避免了待调度无人机调度失败造成的能量浪费。

[0139] 本发明实施例中，根据当前环境状态以及本发明第一方面及其可选方案涉及的方法所确定的Q值表，确定当前协同通信策略，进一步可以理解为，各架待调度无人机自身均存储有对应的Q值表，在Q值表中选择当前环境状态下Q值最大的协同通信策略为当前协同通信策略。

[0140] 本发明实施例中，执行当前协同通信策略中当前待调度无人机所对应的策略，进

一步可以理解为，当前协同通信策略内包括N个通信策略，其中的一个通信策略对应当前待调度无人机，找到N个通信策略中当前待调度无人机对应位置的通信策略并执行。其中，当前待调度无人机所对应的策略执行的结果可以是当前待调度无人机在目标频段通信(即，发送信标信号)或保持静默。

[0141] 图6是本发明一实施例中多无人机协同通信调度的Q值表的学习装置的模块示意图一。

[0142] 请参考图6，一种多无人机协同通信调度的Q值表的学习装置，应用于N架无人机中任意之一的当前无人机，其中的N为大于或等于2的整数，该学习装置，包括：

[0143] 初始Q值表确定模块21，用于确定当前无人机的初始的Q值表；Q值表记载了对应的无人机在每一种环境状态下，N架无人机执行不同的协同通信策略对应的Q值；其中，不同的环境状态表征了目标频段的信号的不同受干扰情况；每个协同通信策略表征了N架无人机中各架无人机的通信策略的一种组合；

[0144] Q值表获取模块22，获取N架无人机中除当前无人机之外其余无人机在第K个时隙的Q值表；

[0145] 更新模块23，用于在N架无人机受控执行任意第K个时隙的第一协同通信策略之后，用于通过Q学习的方式更新当前无人机在第K个时隙的Q值表中的Q值，得到第K+1个时隙的Q值表；第K+1个时隙的Q值表是根据第一协同通信策略、第K个时隙的第一环境状态、第K+1个时隙的第二环境状态，以及其余无人机在第K个时隙的Q值表更新的。

[0146] 图7是本发明一实施例中多无人机协同通信调度的Q值表的学习装置的模块示意图二。

[0147] 请参考图7，更新模块23，包括：

[0148] 反馈单元2301，用于确定当前无人机在第K个时隙对应的调度回报值；调度回报值表征了当前无人机执行第一协同通信策略中相应策略后是否成功通信；

[0149] 第二协同通信策略确定单元2302，用于根据第K个时隙的N张Q值表以及第二环境状态，在所有协同通信策略中确定第二协同通信策略；其中，第二协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的；第二协同通信策略的Q值统计值是针对于第K个时隙的N张Q值表中第二协同通信策略与第二环境状态所对应的N个Q值进行统计而得到的；

[0150] 第一计算单元2303，用于更新当前无人机在第K个时隙的Q值表中的Q值，得到当前无人机的第K+1个时隙对应的Q值表。

[0151] 可选的，所述当前无人机的第K个时隙对应的Q值表中更新的Q值表征为：

[0152] $$Q_n^K(s_K,a)=Q_n^K(s_K,a)+\lambda\left[r_n+\gamma Q_n^K\left(s_{K+1},a^*\right)-Q_n^K(s_K,a)\right]$$；

[0153] 其中，n为当前无人机的编号，Q_n^K为当前无人机在第K个时隙的Q值表，$\lambda\in(0,1)$为学习率；$\gamma\in(0,1)$为算法折扣因子；r_n为调度回报值；a为第一协同通信策略，s_K为第一环境状态，s_{K+1}为第二环境状态，a^*为第二协同通信策略。

[0154] 可选的，第二协同通信策略的Q值统计值为第K个时隙的N张Q值表中第二协同通信策略与第二环境状态所对应的N个Q值之和。

[0155] 图8是本发明一实施例中多无人机协同通信调度的Q值表的学习装置的模块示意图三。

[0156] 请参考图8,该装置还包括第一协同通信策略确定模块24,用于:

[0157] 以概率ε在两个策略确定方式中选择第一策略确定方式后,利用第一策略确定方式确定第一协同通信策略;或者:

[0158] 以概率1-ε,在两个策略确定方式中选择第二策略确定方式后,利用第二策略确定方式确定第一协同通信策略;

[0159] 利用第一策略确定方式确定第一协同通信策略包括:

[0160] 随机确定N架无人机的通信策略的一种组合作为第一协同通信策略;

[0161] 利用第二策略确定方式确定第一协同通信策略,包括:

[0162] 根据第K-1个时隙的N张Q值表以及第一环境状态,在所有协同通信策略中确定第一协同通信策略;第一协同通信策略的Q值统计值是对应时隙所有协同通信策略中最大的;第一协同通信策略的Q值统计值是针对于第K-1个时隙的N张Q值表中第一协同通信策略与第一环境状态所对应的N个Q值进行统计而得到的。

[0163] 以上方案中,通过学习确定的当前无人机的Q值表记载了对应的无人机在每一种环境状态下,N架无人机执行不同的协同通信策略对应的Q值,相对于现有技术,当前无人机可根据对应的已通过学习确定的Q值表独立作出决策;由于不同的环境状态表征了目标频段的信号的不同受干扰情况,每个协同通信策略表征了N架无人机中各架无人机的通信策略的一种组合,进而,当前无人机根据已确定的Q值表作出决策时,已经考虑了当前无人机的飞行环境(目标频段的信号受干扰情况)以及其余无人机的状态(其余无人机的通信策略)。相对于现有技术,N架无人机中任意之一的当前无人机确定通信策略时均增加了作决策时考虑的因素,故而,以上方案涉及的多无人机协同通信调度的Q值表的学习装置,能够提高N架无人机整体的调度成功率。

[0164] 图9是本发明一实施例中电子设备的构造示意图。

[0165] 请参考图9,一种电子设备,包括处理器31与存储器33,

[0166] 存储器33,用于存储代码和相关数据;

[0167] 处理器31,用于执行存储器33中的代码用以实现本发明第一方面及其可选方案涉及的方法。

[0168] 处理器31能够通过总线32与存储器33通讯。

[0169] 本发明一实施例还提供了一种存储介质,其上存储有计算机程序,该程序被处理器31执行时实现本发明第一方面及其可选方案涉及的方法。

[0170] 最后应说明的是:以上各实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述各实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分或者全部技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的范围。

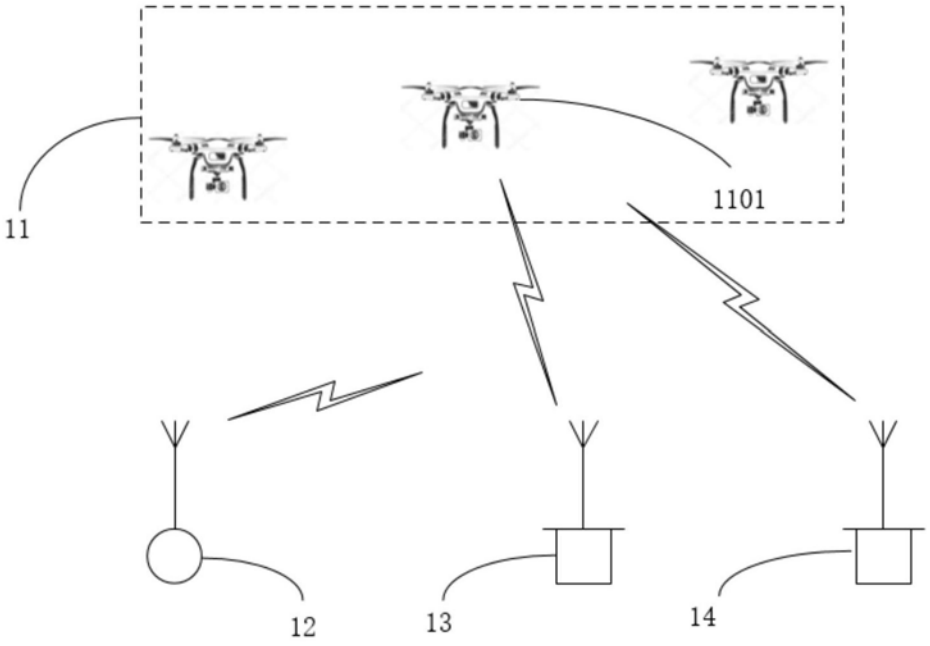

11
1101
12
13
14

图1

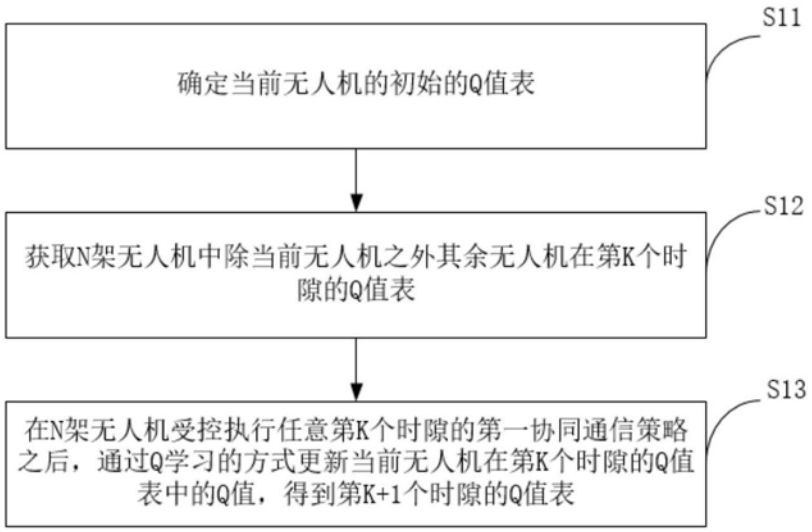

S11
确定当前无人机的初始的Q值表
S12
获取N架无人机中除当前无人机之外其余无人机在第K个时隙的Q值表
S13
在N架无人机受控执行任意第K个时隙的第一协同通信策略之后，通过Q学习的方式更新当前无人机在第K个时隙的Q值表中的Q值，得到第K+1个时隙的Q值表

图2

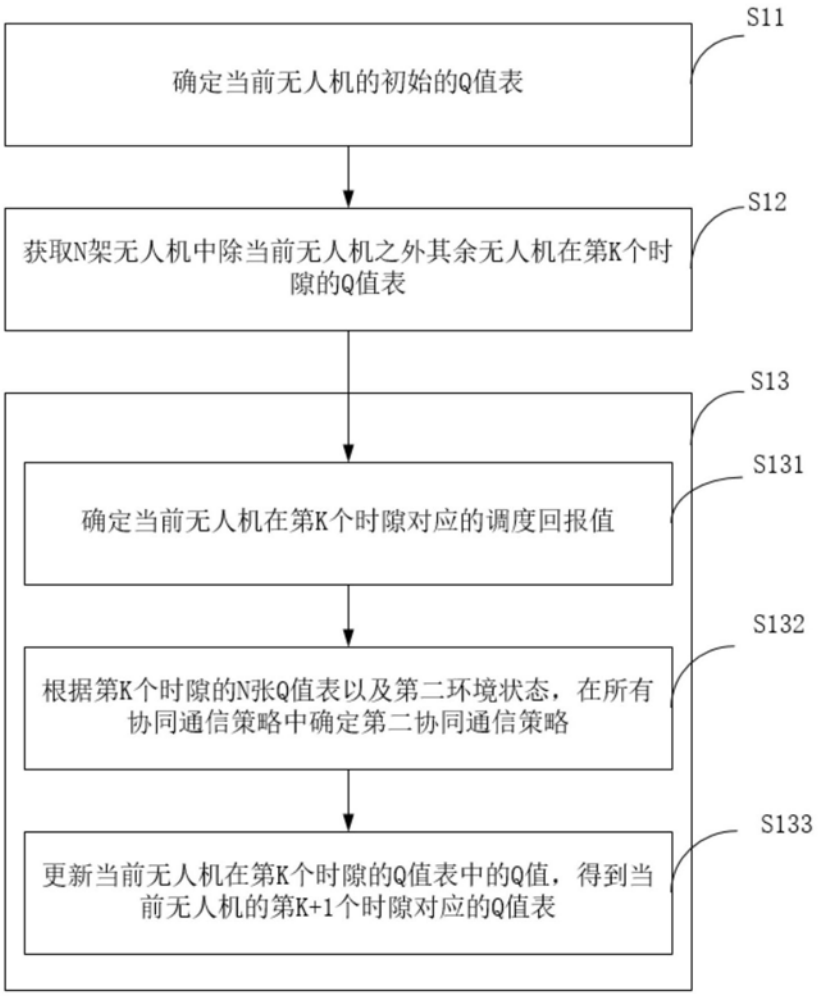

S11
确定当前无人机的初始的Q值表
S12
获取N架无人机中除当前无人机之外其余无人机在第K个时隙的Q值表
S13
S131
确定当前无人机在第K个时隙对应的调度回报值
S132
根据第K个时隙的N张Q值表以及第二环境状态，在所有协同通信策略中确定第二协同通信策略
S133
更新当前无人机在第K个时隙的Q值表中的Q值，得到当前无人机的第K+1个时隙对应的Q值表

图3

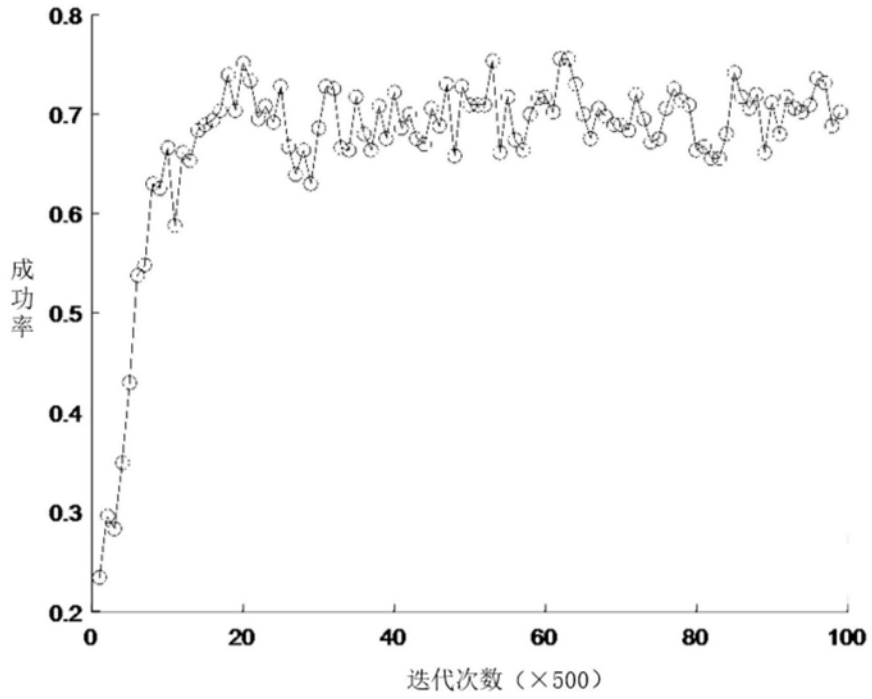

0.8
0.7
0.6
0.5
0.4
0.3
0.2
成功率
0
20
40
60
80
100
迭代次数（×500）

图4

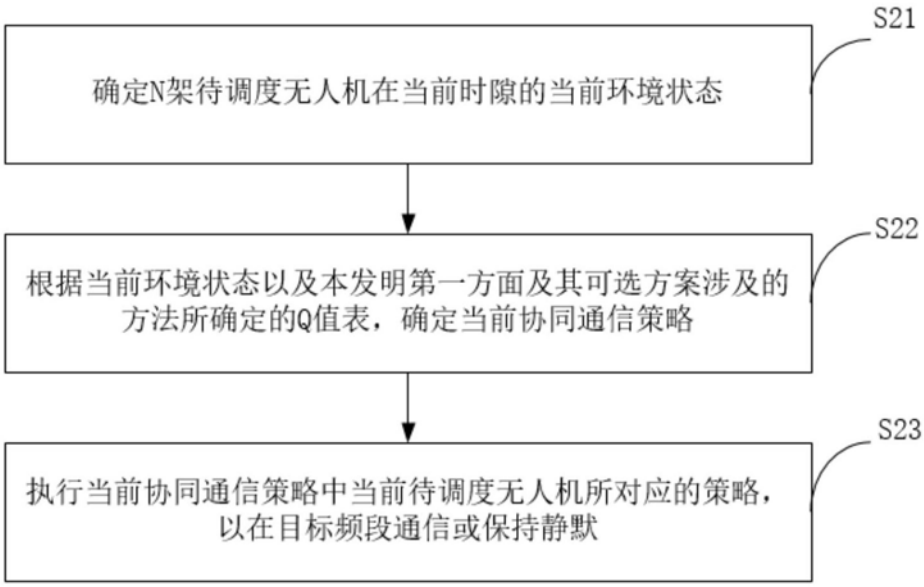

S21
确定N架待调度无人机在当前时隙的当前环境状态
S22
根据当前环境状态以及本发明第一方面及其可选方案涉及的方法所确定的Q值表，确定当前协同通信策略
S23
执行当前协同通信策略中当前待调度无人机所对应的策略，以在目标频段通信或保持静默

图5

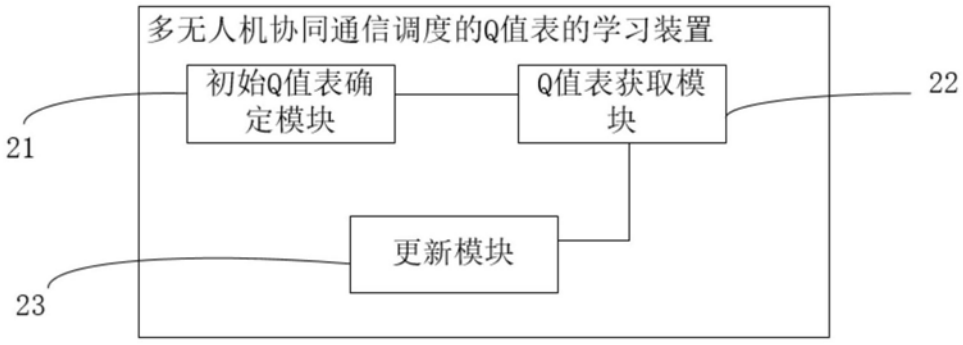

多无人机协同通信调度的Q值表的学习装置
初始Q值表确定模块
Q值表获取模块
更新模块
21
22
23

图6

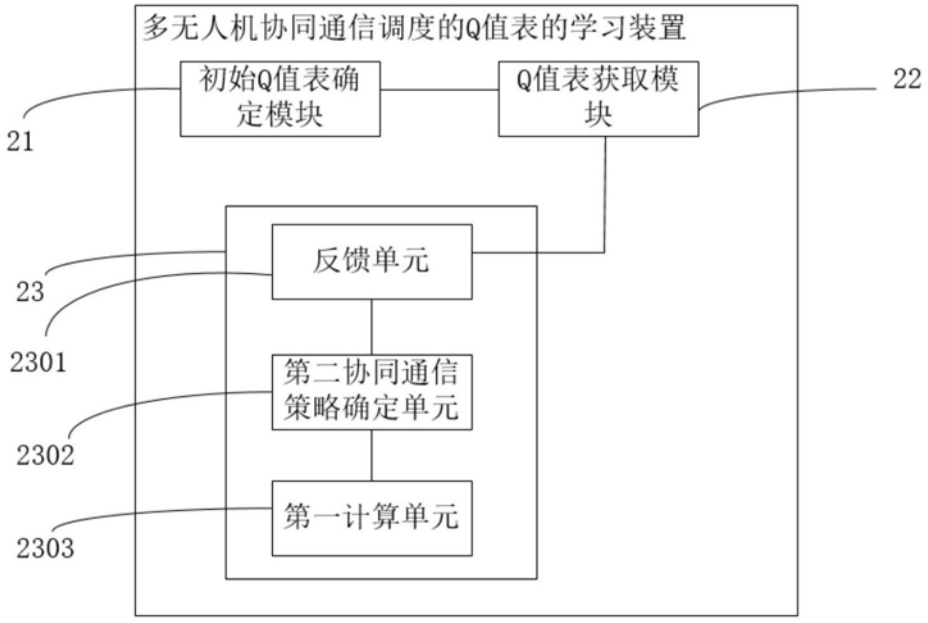

多无人机协同通信调度的Q值表的学习装置
初始Q值表确定模块
Q值表获取模块
反馈单元
第二协同通信策略确定单元
第一计算单元
21
22
23
2301
2302
2303

图7

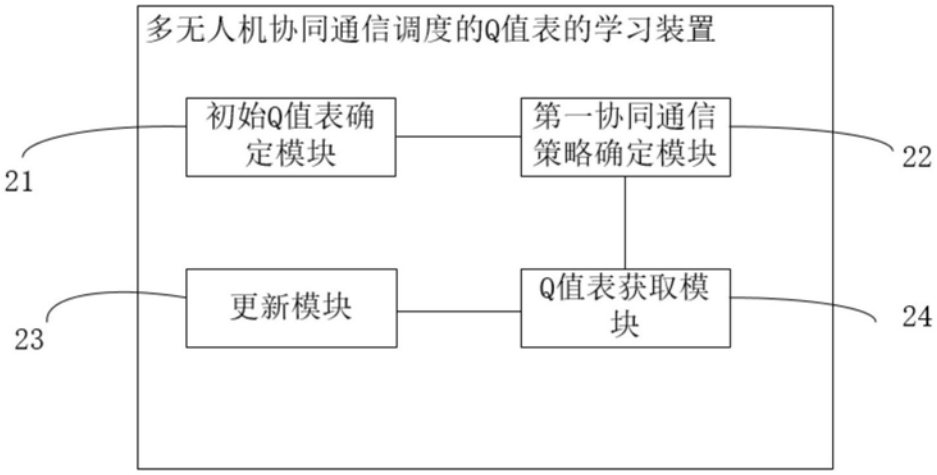

多无人机协同通信调度的Q值表的学习装置
初始Q值表确定模块
第一协同通信策略确定模块
更新模块
Q值表获取模块
21
22
23
24

图8

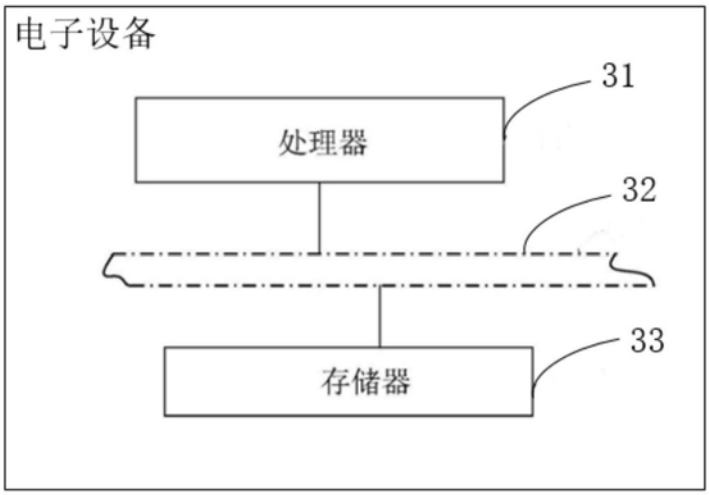

电子设备
处理器
存储器
31
32
33

图9