



(12) 发明专利

(10) 授权公告号 CN 115048859 B

(45) 授权公告日 2025. 05. 30

(21) 申请号 202210577807.6

G06F 18/214 (2023.01)

(22) 申请日 2022.05.25

G06F 18/20 (2023.01)

(65) 同一申请的已公布的文献号

G06F 17/16 (2006.01)

申请公布号 CN 115048859 A

G06F 111/04 (2020.01)

(43) 申请公布日 2022.09.13

(56) 对比文件

(73) 专利权人 青岛科技大学

CN 109814509 A, 2019.05.28

地址 266000 山东省青岛市崂山区松岭路
99号

CN 110674953 A, 2020.01.10

(72) 发明人 李林 陈泽鹏 刘洪果 李华清
尹凤福

审查员 吴海旋

(74) 专利代理机构 青岛中天汇智知识产权代理
有限公司 37241

专利代理师 许莉

(51) Int. Cl.

G06F 30/27 (2020.01)

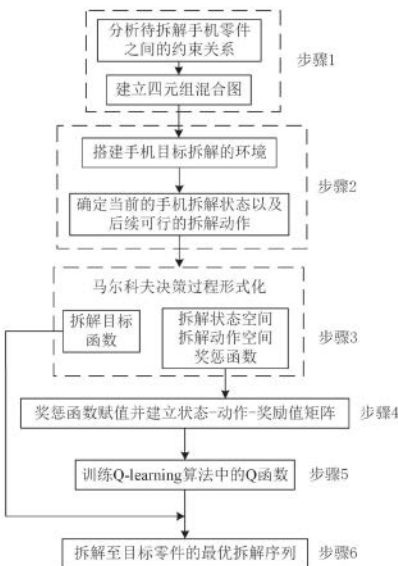
权利要求书2页 说明书9页 附图4页

(54) 发明名称

一种基于强化学习的废旧手机目标拆解序列优化方法

(57) 摘要

本发明公开了一种基于强化学习的废旧手机目标拆解序列优化方法,首先分析待拆解手机零件之间的约束关系,建立四元组混合图,在此基础上搭建手机目标拆解的环境,确定当前的手机拆解状态以及后续可行的拆解动作,然后将废旧手机目标拆解序列问题用马尔科夫决策过程形式化,在此基础上对奖惩函数赋值并建立状态-动作-奖励值矩阵,利用状态-动作-奖励值矩阵对Q-learning算法中的Q函数进行训练,最终利用训练后的Q函数和形式化后的拆解目标函数进行搜寻,得到拆解至目标零件的最优拆解序列。本发明以四元组混合图为基础搭建了更容易编译的强化学习环境,并且将无模型的强化学习算法用于废旧手机目标拆解序列问题,降低了模型构建的困难。



1. 一种基于强化学习的废旧手机目标拆解序列优化方法,其特征在于包含以下步骤:

步骤1、分析待拆解手机零件之间的约束关系,建立四元组混合图 $G=(V,E,D,C)$,四元组混合图包含的信息,具体如下:拆卸的零部件 V 、两个拆卸单元之间的连接关系 E 、两个拆卸单元之间的强物理约束关系 D 、零部件之间不相互连接但存在优先关系 C ,在四元组混合图中, E 用无向边表示, D 用有向边表示, C 用带有箭头的虚线表示;

步骤2、利用步骤1建立的四元组混合图搭建手机目标拆解的环境,确定当前的手机拆解状态以及后续可行的拆解动作,其中搭建手机目标拆解的强化学习环境,具体如下:将待拆解手机的目标零件问题设定为闯关游戏问题,利用建立的四元组混合图中的约束关系表达强化学习环境的约束关系,即将手机拆解混合图中待拆解手机内部零件的约束关系转化为游戏关卡之间的约束,当零件A对零件B为强物理约束关系时,需要先通过关卡A后才能开启关卡B;当零件A对零件B不相互连接但存在优先关系时,需要先通过关卡A后才能开启关卡B;当零件A对零件B为连接关系时,则关卡A与关卡B之间不存在先后关系;将待拆解手机的目标零件设为目标关卡,要想得到最大的奖励,必须到达设定的目标关卡;

步骤3、将废旧手机目标拆解序列问题用马尔科夫决策过程形式化,具体包括:拆解状态空间、拆解动作空间、奖惩函数和拆解目标函数;

步骤4、设定待拆解手机的目标零件,并根据步骤3形式化后的拆解状态空间、拆解动作空间对奖惩函数赋值,建立状态-动作-奖励值矩阵;

步骤5、利用步骤4建立的状态-动作-奖励值矩阵对Q-learning算法中的Q函数进行训练;

步骤6、利用步骤5训练后的Q函数和步骤3形式化后的拆解目标函数进行搜寻,得到拆解至目标零件的最优拆解序列。

2. 根据权利要求1所述的一种基于强化学习的废旧手机目标拆解序列优化方法,其特征在于,步骤3所述的将废旧手机目标拆解序列问题用马尔科夫决策过程形式化表示,具体包括:

设置拆解状态空间 S :

$$S=[S_0, S_1, S_2, S_3, S_4 \cdots S_n] \quad (1)$$

其中 $S_n, n=0, 1, 2 \cdots$ 表示拆解至零件 n 时,待拆解手机的状态;

设置拆解动作空间 D :

$$D=[D_0, D_1, D_2, D_3, D_4 \cdots D_n] \quad (2)$$

其中 $D_n, n=0, 1, 2 \cdots$ 表示进行拆解零件 n 的动作;

设置奖惩函数 R :

$$R=[q, w, e] \quad (3)$$

其中 q 表示零件存在约束无法拆解, w 表示零件可以拆解但未拆解至目标零件, e 表示零件可以拆解并拆解至目标零件。

3. 根据权利要求1所述的一种基于强化学习的废旧手机目标拆解序列优化方法,其特征在于,步骤4所述的奖惩函数 R 赋值规则,具体包括:

$$R = \begin{cases} -1, & \text{零件存在约束无法拆解} \\ 0, & \text{未拆解至目标零件} \\ 100, & \text{拆解至目标零件} \end{cases} \quad (4)$$

当零件存在约束时,该状态下无法拆解该零件,若强行拆解则会给予一个值为负数的惩罚,赋值为-1;当未拆解至目标零件时,拆解任意零件不会给予奖励或者惩罚,赋值为0;当零件拆解至目标零件时,该状态给予一个值为正数的奖励,赋值为100。

一种基于强化学习的废旧手机目标拆解序列优化方法

技术领域

[0001] 本发明涉及废旧电子产品拆解工艺领域,具体涉及一种基于强化学习的废旧手机目标拆解序列优化方法。

背景技术

[0002] 随着科学技术的发展和人们生活水平提高,智能手机更新换代速度逐渐加快,产生了大量急需妥善处理的废旧手机。目前手机拆解工厂的工作人员通常根据以往的经验来确定手机拆解序列,且不同手机内部零件约束关系不同导致拆解序列混乱,使拆解效率低下。

[0003] 现有的废旧电子产品拆解工艺领域,专利公布号为CN113477679A的专利是在进行人工拆解或机械设备拆解之前对大批量废旧手机拆解工艺进行规划并生成可行的拆解序列和工艺,但并没有对序列进行优化,不能得到最优或次优的拆解序列;专利公布号为CN113177313A的专利是通过规划手机拆解流水线、制定手机相似度判定方式、完善整体手机拆解流水线的方法,从而达到实现各型号智能手机自动识别、分类,实现不同类型手机同时拆解的目的,但并不涉及具体单个手机的拆解序列;专利公布号为CN113283616A的专利构建了零件回收综合评价体系和评估回收价值,并结合回收评价采用拓扑排序的方法确定完全拆解序列,但将产品完全拆解则会拆解较多低价值的零件,降低了拆解效率。与现有专利技术相比,本发明以高价值零件为拆解目标,拆解至目标零件即结束,与产品完全拆解有本质区别。

[0004] 针对拆解序列规划问题,目前的研究大多数利用启发式搜索规划算法,如遗传算法和蚁群算法,但相比于无模型的强化学习算法,启发式搜索规划算法需要构建大量具体问题模型,导致算法计算量较大、通用性不足。

发明内容

[0005] 本发明的主要目的在于克服现有技术的缺点与不足,提出一种基于强化学习的废旧手机目标拆解序列优化方法,利用无模型的强化学习算法对待拆解手机的目标拆解序列进行智能决策,找到最优的拆解序列。

[0006] 一种基于强化学习的废旧手机目标拆解序列优化方法,包括以下步骤:

[0007] 步骤1、分析待拆解手机零件之间的约束关系,建立四元组混合图;

[0008] 步骤2、利用步骤1建立的四元组混合图搭建手机目标拆解的环境,确定当前的手机拆解状态以及后续可行的拆解动作;

[0009] 步骤3、将废旧手机目标拆解序列问题用马尔科夫决策过程形式化,具体包括:拆解状态空间、拆解动作空间、奖惩函数和拆解目标函数;

[0010] 步骤4、设定待拆解手机的目标零件,并根据步骤3形式化后的拆解状态空间、拆解动作空间对奖惩函数赋值,建立状态-动作-奖励值矩阵;

[0011] 步骤5、利用步骤4建立的状态-动作-奖励值矩阵对Q-learning算法中的Q函数进

行训练；

[0012] 步骤6、利用步骤5训练后的Q函数和步骤3形式化后的拆解目标函数进行搜寻,得到拆解至目标零件的最优拆解序列。

[0013] 所述步骤1建立的四元组混合图,具体如下:

[0014] 建立 $G=(V,E,D,C)$ 的四元组混合图,包含的信息为拆卸的零部件V、两个拆卸单元之间的连接关系E、两个拆卸单元之间的强物理约束关系D、零部件之间不相互连接但存在优先关系C,在四元组混合图中,E用无向边表示,D用有向边表示,C用带有箭头的虚线表示;

[0015] 当零件A对零件B为连接关系($A-B$)时,零件A和零件B可以随意拆解;

[0016] 当零件A对零件B为强物理约束关系($A \rightarrow B$)时,只能先拆卸零件A再拆卸零件B;

[0017] 当零件A对零件B为不相互连接但存在优先关系($A \rightarrow B$)时,只能先拆卸零件A再拆卸零件B。

[0018] 所述步骤2搭建手机拆解的强化学习环境,具体如下:

[0019] 将手机拆解四元组混合图转化为手机拆解的强化学习环境,将待拆解手机的目标零件问题设定为闯关游戏问题(强化学习环境),利用建立的四元组混合图中的约束关系表达强化学习环境的约束关系,即将手机拆解混合图中待拆解手机内部零件的约束关系转化为游戏关卡之间的约束,当零件A对零件B为强物理约束关系时,需要先通过关卡A后才能开启关卡B;当零件A对零件B不相互连接但存在优先关系时,需要先通过关卡A后才能开启关卡B;当零件A对零件B为连接关系时,则关卡A与关卡B之间不存在先后关系;将待拆解手机的目标零件设为目标关卡,要想得到最大的奖励,必须到达设定的目标关卡。

[0020] 所述步骤3的废旧手机目标拆解序列问题形式化,具体包括:

[0021] 废旧手机目标拆解序列问题是在待拆解手机目前的拆解状态下,为拆解目标零件,对后续将要拆解的零件做出决策,决策出拆解至目标零件前,所需拆解最少的非目标零件的序列,该问题属于一种序贯决策问题,所以本发明采用马尔科夫决策过程形式化;

[0022] 设置拆解状态空间S:

$$[0023] \quad S=[S_0, S_1, S_2, S_3, S_4 \cdots S_n] \quad (1)$$

[0024] 其中 $S_n, n=0, 1, 2 \cdots n$ 表示拆解至零件n时,待拆解手机的状态;

[0025] 设置拆解动作空间D:

$$[0026] \quad D=[D_0, D_1, D_2, D_3, D_4 \cdots D_n] \quad (2)$$

[0027] 其中 $D_n, n=0, 1, 2 \cdots$ 表示进行拆解零件n的动作;

[0028] 设置奖惩函数R:

$$[0029] \quad R=[q, w, e] \quad (3)$$

[0030] 其中q表示零件存在约束无法拆解,w表示零件可以拆解但未拆解至目标零件,e表示零件可以拆解并拆解至目标零件。

[0031] 拆解目标函数:对于废旧手机目标拆解序列规划问题,强化学习算法通过用值函数定义折扣累计回报,值函数越大,累计回报越大,越符合拆解的目标序列,折扣累计回报如下:

$$[0032] \quad q_{\pi}(S, D)=E_{\pi}\left[G_n \mid S_n=S, D_n=D\right]=E_{\pi}\left[\sum_{k=0}^{\infty} \gamma^k R_{n+k+1} \mid S_n=S, D_n=D\right] \quad (4)$$

$$[0033] \quad \pi=S_0, D_0, \gamma_0, S_1, D_1, \gamma_1 \cdots \quad (5)$$

[0034] 其中 q 为状态-动作值函数, E 为期望, S 为状态, D 为动作, n 为零件代号, γ 为折扣因子, k 为时间步长, π 为从初始状态 S_0 开始,采取某一动作 D ,得到奖惩值 R_{n+k+1} ,直至终止状态。

[0035] 所述步骤4的奖惩函数 R 赋值规则,具体包括:

$$[0036] \quad R = \begin{cases} -1, & \text{零件存在约束无法拆解} \\ 0, & \text{未拆解至目标零件} \\ 100, & \text{拆解至目标零件} \end{cases} \quad (6)$$

[0037] 当零件存在约束时,该状态下无法拆解该零件,若强行拆解则会给予一个值为负数的惩罚,赋值为-1;当未拆解至目标零件时,拆解任意零件不会给予奖励或者惩罚,赋值为0;当零件拆解至目标零件时,该状态给予一个值为正数的奖励,赋值为100。

[0038] 所述步骤4建立的状态-动作-奖励值矩阵,具体包括:

[0039] 建立一个 $n \times n$ 矩阵($n=1,2,3,4\cdots$),矩阵的行表示接下来将要进行拆解的零件;矩阵的列表示手机拆解所处的状态;矩阵内的元素表示在某一状态下进行某一动作所获得的奖励值;例如,矩阵第三行为零件A,第四列为执行动作拆解零件B,所对应的元素为-1表示为手机拆解至零件A状态下将要拆解零件B,但零件B存在约束无法拆解零件B,则给予一个值为-1的惩罚;当矩阵第二行为零件C,第四列为执行动作拆解零件D,所对应的元素为0,表示为手机拆解至零件C状态下将要拆解零件D,零件D在拆解完零件C后无约束,可以进行拆解零件D,但零件D不是目标零件,则不给予惩罚和奖励;当矩阵的第一行为零件E,第三列为执行动作拆解零件F,所对应的元素为100表示为手机拆解至零件E状态下将要拆解零件F,零件F在拆解完零件E后无约束,可以进行拆解零件F,而且零件F为目标零件,则给予一个值为100的奖励。

[0040] 本发明与现有技术相比,具有如下优点和有益效果:

[0041] 1) 本发明以四元组混合图为基础对强化学习的环境进行搭建,使强化学习环境更容易编译。2) 本发明将无模型的强化学习算法用于手机拆解问题中,解决了以往拆解问题建模困难和通用性差的问题,决策出拆解至目标零件前,所需拆解最少的非目标零件的序列。

附图说明

[0042] 图1为基于强化学习的废旧手机目标拆解序列优化方法流程图;

[0043] 图2为苹果7零件约束关系图;

[0044] 图3为以苹果7后置摄像头为目标零件的最优序列生成结果图;

[0045] 图4为以苹果7主板为目标零件的最优序列生成结果图;

[0046] 图5为小米5零件约束关系图;

[0047] 图6为以小米5后置摄像头为目标零件的最优序列生成结果图;

[0048] 图7为以小米5主板为目标零件的最优序列生成结果图。

具体实施方式

[0049] 为使本发明的目的、技术方案及优点更加清楚明白,以下通过实施例,并结合附图对本发明进行进一步详细说明,但不应该理解为本发明上述主题范围仅限于下述实施例。在不脱离本发明上述技术思想的情况下,根据本领域普通技术知识和惯用手段,做出各种

替换和变更,均应包括在本发明范围内。

[0050] 参照图1,一种基于强化学习的废旧手机目标拆解序列优化方法,包括以下步骤:

[0051] 步骤1、分析待拆解手机零件之间的约束关系,建立四元组混合图;

[0052] 步骤2、利用步骤1建立的四元组混合图搭建手机目标拆解的环境,确定当前的手机拆解状态以及后续可行的拆解动作;

[0053] 步骤3、将废旧手机目标拆解序列问题用马尔科夫决策过程形式化,具体包括:拆解状态空间、拆解动作空间、奖惩函数和拆解目标函数;

[0054] 步骤4、设定待拆解手机的目标零件,根据步骤2和步骤3形式化后的拆解状态空间、拆解动作空间对奖惩函数赋值,建立状态-动作-奖励值矩阵;

[0055] 步骤5、利用步骤4建立的状态-动作-奖励值矩阵对Q-learning算法中的Q函数进行训练;

[0056] 步骤6、利用步骤5训练后的Q函数和步骤3形式化后的拆解目标函数进行搜寻,得到拆解至目标零件的最优拆解序列。

[0057] 所述步骤1建立的四元组混合图,具体如下:

[0058] 建立 $G = (V, E, D, C)$ 的四元组混合图,包含的信息为拆卸的零部件V、两个拆卸单元之间的连接关系E、两个拆卸单元之间的强物理约束关系D、零部件之间不相互连接但存在优先关系C,在四元组混合图中,E用无向边表示,D用有向边表示,C用带有箭头的虚线表示;

[0059] 当零件A对零件B为连接关系($A-B$)时,零件A和零件B可以随意拆解;

[0060] 当零件A对零件B为强物理约束关系($A \rightarrow B$)时,只能先拆卸零件A再拆卸零件B;

[0061] 当零件A对零件B为不相互连接但存在优先关系($A \rightarrow B$)时,只能先拆卸零件A再拆卸零件B。

[0062] 所述步骤2搭建手机拆解的强化学习环境,具体如下:

[0063] 将手机拆解四元组混合图转化为手机拆解的强化学习环境,将待拆解手机的目标零件问题设定为闯关游戏问题(强化学习环境),利用建立的四元组混合图中的约束关系表达强化学习环境的约束关系,即将手机拆解混合图中待拆解手机内部零件的约束关系转化为游戏关卡之间的约束,当零件A对零件B为强物理约束关系时,需要先通过关卡A后才能开启关卡B;当零件A对零件B不相互连接但存在优先关系时,需要先通过关卡A后才能开启关卡B;当零件A对零件B为连接关系时,则关卡A与关卡B之间不存在先后关系;将待拆解手机的目标零件设为目标关卡,要想得到最大的奖励,必须到达设定的目标关卡。

[0064] 所述步骤3的废旧手机目标拆解序列问题形式化,具体包括:

[0065] 废旧手机目标拆解序列问题是在待拆解手机目前的拆解状态下,为拆解目标零件,对后续将要拆解的零件做出决策,决策出拆解至目标零件前,所需拆解最少的非目标零件的序列,该问题属于一种序贯决策问题,所以本发明采用马尔科夫决策过程形式化;

[0066] 设置拆解状态空间S:

$$S = [S_0, S_1, S_2, S_3, S_4 \cdots S_n] \quad (7)$$

[0068] 其中 $S_n, n=0, 1, 2 \cdots n$ 表示拆解至零件n时,待拆解手机的状态;

[0069] 设置拆解动作空间D:

$$D = [D_0, D_1, D_2, D_3, D_4 \cdots D_n] \quad (8)$$

[0071] 其中 $D_n, n=0, 1, 2 \cdots$ 表示进行拆解零件n的动作;

[0072] 设置奖惩函数R:

$$[0073] \quad R=[q, w, e] \quad (9)$$

[0074] 其中q表示零件存在约束无法拆解,w表示零件可以拆解但未拆解至目标零件,e表示零件可以拆解并拆解至目标零件。

[0075] 拆解目标函数:对于废旧手机目标拆解序列规划问题,强化学习算法通过用值函数定义折扣累计回报,值函数越大,累计回报越大,越符合拆解的目标序列,折扣累计回报如下:

$$[0076] \quad q_{\pi}(S, D) = E_{\pi} [G_n | S_n = S, D_n = D] = E_{\pi} \left[\sum_{k=0}^{\infty} \gamma^k R_{n+k+1} | S_n = S, D_n = D \right] \quad (10)$$

$$[0077] \quad \pi = S_0, D_0, \gamma_0, S_1, D_1, \gamma_1 \dots \quad (11)$$

[0078] 其中q为状态-动作值函数,E为期望,S为状态,D为动作,n为零件代号, γ 为折扣因子,k为时间步长, π 为从初始状态 S_0 开始,采取某一动作D,得到奖惩值 R_{n+k+1} ,直至终止状态。

[0079] 所述步骤4所述的奖惩函数R,具体包括:

$$[0080] \quad R = \begin{cases} -1, & \text{零件存在约束无法拆解} \\ 0, & \text{未拆解至目标零件} \\ 100, & \text{拆解至目标零件} \end{cases} \quad (12)$$

[0081] 当零件存在约束时,该状态下无法拆解该零件,若强行拆解则会给予一个值为负数的惩罚,赋值为-1;当未拆解至目标零件时,拆解任意零件不会给予奖励或者惩罚,赋值为0;当零件拆解至目标零件时,该状态给予一个值为正数的奖励,赋值为100。

[0082] 所述步骤4建立的状态-动作-奖励值矩阵,具体包括:

[0083] 建立一个 $n \times n$ 矩阵($n=1, 2, 3, 4 \dots$),矩阵的行表示接下来将要进行拆解的零件;矩阵的列表示手机拆解所处的状态;矩阵内的元素表示在某一状态下进行某一动作所获得的奖励值;例如,矩阵第三行为零件A,第四列为执行动作拆解零件B,所对应的元素为-1表示为手机拆解至零件A状态下将要拆解零件B,但零件B存在约束无法拆解零件B,则给予一个值为-1的惩罚;当矩阵第二行为零件C,第四列为执行动作拆解零件D,所对应的元素为0,表示为手机拆解至零件C状态下将要拆解零件D,零件D在拆解完零件C后无约束,可以进行拆解零件D,但零件D不是目标零件,则不给予惩罚和奖励;当矩阵的第一行为零件E,第三列为执行动作拆解零件F,所对应的元素为100表示为手机拆解至零件E状态下将要拆解零件F,零件F在拆解完零件E后无约束,可以进行拆解零件F,而且零件F为目标零件,则给予一个值为100的奖励。

[0084] 目前手机市场两大主流操作系统为IOS系统和Android系统,故选取IOS系统的苹果7和Android系统的小米5作为本发明的实施例,对两种品牌手机进行目标零件的拆解序列优化;目前市场条件下具有高回收价值的手机零件主要有摄像头和主板,故本实施例选取主板和摄像头作为拆解的目标零件。

[0085] 实施例1:分析苹果7内部零件相关信息,苹果7包含的主要零件有屏幕、震动模块、扬声器、摄像头、电池、主板、lightning接口和后盖,将零件进行编号,见表1。

[0086] 表1苹果7手机零件代号表

[0087]

零件代号	0	1	2	3	4	5	6	7
零件名称	屏幕	震动模块	扬声器	主板	摄像头	电池	Lightning接口	后盖

[0088] 分析内部零件之间的约束关系。屏幕对震动模块、扬声器和摄像头具有强约束关系,震动模块对电池和扬声器具有强约束关系,扬声器对主板有强约束关系,主板对lightning接口具有强约束关系,lightning接口对后盖有强约束关系;主板对后盖为不相互连接但存在优先关系;电池对后盖也为不相互连接但存在优先关系;强约束关系用有向边($\rightarrow$)表示,不相互连接但存在优先关系用带有箭头的虚线($\dashrightarrow$)表示。综上分析,建立苹果7零件约束关系图,如图2所示。

[0089] 构建强化学习的环境时将手机拆解问题假设为闯关问题,之后将四元组混合图转化成强化学习的环境(闯关问题),具体如下:屏幕对震动模块、扬声器和摄像头具有强约束关系,即需要通过屏幕所在的关卡才能通往震动模块、扬声器和摄像头所在的关卡;震动模块对电池和扬声器具有强约束关系,即需要通过震动模块所在的关卡才能通往电池和扬声器所在的关卡;扬声器对主板有强约束关系,即需要通过扬声器所在的关卡才能通往主板所在的关卡;主板对lightning接口具有强约束关系,即需要通过主板所在的关卡才能通过lightning接口所在的关卡;主板对后盖为不相互连接但存在优先关系,也需要通过主板所在的关卡才能通过后盖所在的关卡;电池对后盖也为不相互连接但存在优先关系,也需要通过主板所在的关卡才能通过后盖所在的关卡。

[0090] 以摄像头作为拆解的目标零件时,目标关卡也就变成摄像头所在的关卡,所述的奖惩函数的目标零件也为摄像头,即当拆解至摄像头时,获得一个值为100的奖励值;拆解至其他无约束的零件时,既不会得到惩罚也不会得到奖励;拆解存在约束的零件导致无法拆解时,会给予一个值为-1的惩罚。由于目标零件为摄像头,在建立动作状态—奖励值矩阵L时,只需要找出与摄像头存在关联的零件并分析之间的约束关系即可,即屏幕对摄像头存在强约束关系。之后根据约束关系建立矩阵L,表示以摄像头作为目标零件时,不同的状态下拆解不同的苹果7零件所获得的奖励值,矩阵的列表示将要拆解的手机目前的状态,矩阵的行表示将要拆解的手机零件。

[0091] 根据约束关系完成矩阵L,如下所示:

$$[0092] \quad L = \begin{pmatrix} -1 & 0 & -1 & -1 & 100 \\ -1 & -1 & 0 & -1 & 100 \\ -1 & -1 & -1 & 0 & 100 \\ -1 & -1 & -1 & -1 & 100 \\ -1 & 0 & -1 & -1 & 100 \end{pmatrix}$$

[0093] 将Q-learning策略算法中的Q函数进行训练,并利用训练后的Q函数,以电池为目标零件并根据矩阵L所表示的动作、状态和奖励值进行搜寻,找到最大的奖励值,从而得到最优拆解序列,最终以苹果7的摄像头为目标零件生成最优序列为:0—4,屏幕—摄像头,如图3所示。

[0094] 以主板作为拆解的目标零件时,目标关卡也就变成主板所在的关卡,所述的奖惩函数的目标零件也为主板,即当拆解至主板时,获得一个值为100的奖励值;拆解至其他无约束的零件时,既不会得到惩罚也不会得到奖励;拆解存在约束的零件导致无法拆解时,会给予一个值为-1的惩罚。由于目标零件为主板,在建立动作状态—奖励值矩阵R时,只需要找出与主板存在关联的零件并分析之间的约束关系即可,即扬声器对主板存在强约束关系,震动模块对扬声器存在强约束关系,屏幕对震动模块和扬声器存在强约束关系。之后根

据约束关系建立矩阵R,表示以主板作为目标零件时,不同的状态下拆解不同的苹果7零件所获得的奖励值,矩阵的列表示将要拆解的手机目前的状态,矩阵的行表示将要拆解的手机零件。

[0095] 最后根据约束关系完成矩阵R,如下所示:

$$[0096] \quad R = \begin{pmatrix} -1 & 0 & -1 & -1 \\ -1 & -1 & 0 & -1 \\ -1 & -1 & -1 & 100 \\ -1 & -1 & -1 & 100 \end{pmatrix}$$

[0097] 将Q-learning异策略算法中的Q函数进行训练,并利用训练后的Q函数,以主板为目标零件并根据矩阵R所表示的动作、状态和奖励值进行搜寻,找到最大的奖励值,从而得到最优拆解序列,最终以苹果7的主板为目标零件生成最优序列为:0—1—2—3,屏幕—震动模块—扬声器—主板,如图4所示。

[0098] 为表明利用强化学习的序列优化结果,选取两组人员以苹果7的主板为目标零件进行拆解,A组选取三人,在无方法指导下,仅凭经验进行拆解,B组选取三人,在本方法指导下进行拆解;最终A组一号和二号在拆解至目标零件前拆解了四个手机零件,A组三号在拆解至目标零件前拆解五个手机零件,B组的一号、二号和三号在拆解至目标零件前均拆解三个手机零件,如表2所示。

[0099] 表2以苹果7主板为目标零件的拆解序列表

[0100]	A 组	一号	0-1-4-2-3	屏幕-震动模块-摄像头-扬声器-主板
		二号	0-4-1-2-3	屏幕-摄像头-震动模块-扬声器-主板
		三号	0-1-5-2-3	屏幕-震动模块-电池-扬声器-主板
	B 组	一号	0-1-2-3	屏幕-震动模块-扬声器-主板
		二号	0-1-2-3	屏幕-震动模块-扬声器-主板
		三号	0-1-2-3	屏幕-震动模块-扬声器-主板

[0101] 结果表明,相比于以往的人工经验拆解,在本方法指导下的拆解,在拆解至目标零件前,所需要拆解的零件数量有所减少。

[0102] 实施例2:分析小米5内部零件相关信息,小米5包含的主要零件有后盖、主板盖板、电池、主板、后置摄像头盖板、后置摄像头、前置摄像头、尾板盖板、尾板和屏幕,为方便后续研究利用,将零件进行编号,见表3。

[0103] 表3小米5手机零件代号表

[0104]	零件代号	0	1	2	3	4
	零件名称	后盖	主板盖板	电池	主板	后置摄像头盖板
	零件代号	5	6	7	8	9
	零件名称	后置摄像头	前置摄像头	尾板盖板	尾板	屏幕

[0105] 分析内部零件之间的约束关系,后盖对尾板盖板和主板盖板具有强约束关系,主板盖板对电池、前置摄像头、主板和后置摄像头盖板具有强约束关系,尾板盖板对尾板有强

约束关系,后置摄像头盖板对后置摄像头具有强约束关系,主板对屏幕有强约束关系,电池对尾板和屏幕有强约束关系,尾板对屏幕有强约束关系;主板盖板对屏幕为不相互连接但存在优先关系;尾板盖板对屏幕也为不相互连接但存在优先关系,主板与后置摄像头为连接关系;强约束关系用有向边($\rightarrow$)表示,不相互连接但存在优先关系用带有箭头的虚线($\dashrightarrow$)表示,连接关系用直线($—$)表示。综上分析,建立小米5零件约束关系图,如图5所示。

[0106] 构建强化学习的环境时将拆解问题假设为闯关问题,之后将四元组混合图转化成强化学习的环境(闯关问题),具体如下:后盖对尾板盖板和主板盖板具有强约束关系,即需要通过后盖所在的关卡才能通往尾板盖板和主板盖板所在的关卡;主板盖板对电池、前置摄像头、主板和后置摄像头盖板具有强约束关系,即需要通过主板盖板所在的关卡才能通往电池、前置摄像头、主板和后置摄像头所在的关卡;主板对屏幕有强约束关系,即需要通过主板所在的关卡才能通往屏幕所在的关卡;电池对尾板和屏幕有强约束关系,即需要通过电池所在的关卡才能通过尾板和屏幕所在的关卡;尾板对屏幕有强约束关系,即需要通过尾板所在的关卡才能通过屏幕所在的关卡;尾板盖板对屏幕为不相互连接但存在优先关系,也需要通过尾板盖板所在的关卡才能通过屏幕所在的关卡;主板与后置摄像头为连接关系,则两个零件之间无约束关系,故两者所在的关卡没有先后关系。

[0107] 以后置摄像头作为拆解的目标零件时,目标关卡也就变成后置摄像头所在的关卡,所述的奖惩函数的目标零件也为后置摄像头,即拆解至后置摄像头时,获得一个值为100的奖励值;拆解至其他无约束的零件时,既不会得到惩罚也不会得到奖励;拆解存在约束的零件导致无法拆解时,会给予一个值为-1的惩罚。由于目标零件为后置摄像头,在建立动作状态—奖励值矩阵P时,只需要找出与后置摄像头存在关联的零件并分析零件之间的约束关系即可,即后盖对主板盖板存在强约束关系,主板盖板对后置摄像头盖板存在强约束关系,后置摄像头盖板对后置摄像头存在强约束关系。之后根据约束关系建立矩阵P,表示以后置摄像头作为目标零件时,不同的状态下拆解不同的小米5零件所获得的奖励值,矩阵的列表示将要拆解的手机目前的状态,矩阵的行表示将要拆解的手机零件。

[0108] 最后根据约束关系完成矩阵P,如下所示:

$$[0109] \quad P = \begin{pmatrix} -1 & 0 & -1 & -1 & -1 & -1 \\ -1 & -1 & 0 & 0 & 0 & -1 \\ -1 & -1 & -1 & 0 & 0 & -1 \\ -1 & -1 & 0 & -1 & 0 & -1 \\ -1 & -1 & 0 & 0 & -1 & 100 \\ -1 & -1 & 0 & 0 & -1 & 100 \end{pmatrix}$$

[0110] 将Q-learning异策略算法中的Q函数进行训练,并利用训练后的Q函数,以后置摄像头为目标零件并根据矩阵P所表示的动作、状态和奖励值进行搜寻,找到最大的奖励值,从而得到最优拆解序列,最终以小米5的后置摄像头为目标零件生成最优序列为:0—1—4—5,后盖—主板盖板—后置摄像头盖板—后置摄像头,如图6所示。

[0111] 以主板作为拆解的目标零件时,目标关卡也就变成主板所在的关卡,所述的奖惩函数的目标零件也为主板,即当拆解至主板时,获得一个值为100的奖励值;拆解至其他无约束的零件时,既不会得到惩罚也不会得到奖励;拆解存在约束的零件导致无法拆解时,会给予一个值为-1的惩罚。由于目标零件为主板,在建立动作状态—奖励值矩阵U时,只需要

找出与主板存在关联的零件并分析之间的约束关系即可,即主板盖板对主板存在强约束关系,后盖对主板盖板存在强约束关系,后置摄像头对主板存在连接关系。之后根据约束关系建立矩阵U,表示以主板作为目标零件时,在不同的状态下拆解不同的小米5零件所获得的奖励值,矩阵的列表示将要拆解的手机目前的状态,矩阵的行表示将要拆解的手机零件。

[0112] 最后根据约束关系完成矩阵U,如下所示:

[0113]
$$U = \begin{pmatrix} -1 & 0 & -1 & -1 & -1 & -1 \\ -1 & -1 & 0 & 100 & 0 & -1 \\ -1 & -1 & -1 & 100 & 0 & -1 \\ -1 & -1 & 0 & 100 & 0 & -1 \\ -1 & -1 & 0 & 100 & -1 & 0 \\ -1 & -1 & 0 & 100 & -1 & -1 \end{pmatrix}$$

[0114] 将Q-learning异策略算法中的Q函数进行训练,并利用训练后的Q函数,以主板为目标零件并根据矩阵U所表示的动作、状态和奖励值进行搜寻,找到最大的奖励值,从而得到最优拆解序列,最终以小米5的主板为目标零件生成最优序列为:0—1—3,后盖—主板盖板—主板,如图7所示。

[0115] 为表明利用强化学习的序列优化结果,选取两组人员以小米5的主板为目标零件进行拆解,C组选取三人,在无方法指导下,仅凭经验进行拆解,D组选取三人,在本方法指导下进行拆解;最终C组一号在拆解至目标零件前拆解了五个手机零件,C组二号在拆解至目标零件前拆解了四个手机零件,C组三号在拆解至目标零件前拆解了三个手机零件,D组的一号、二号和三号在拆解至目标零件前均拆解两个手机零件,如表4所示。

[0116] 表4以小米5主板为目标零件的拆解序列表

[0117]

C 组	一号	0-7-1-4-5-3	后盖-尾板盖板-主板盖板-后置摄像头盖板-主板	
	二号	0-1-6-4-3	后盖-主板盖板-前置摄像头-后置摄像头盖板-主板	
	三号	0-1-2-3	后盖-主板盖板-电池-主板	
	D 组	一号	0-1-3	后盖-主板盖板-主板
		二号	0-1-3	后盖-主板盖板-主板
		三号	0-1-3	后盖-主板盖板-主板

[0118] 结果表明,相比于以往的人工经验拆解,在本方法指导下的拆解,在拆解至目标零件前,所需要拆解的零件数量有所减少。

[0119] 本发明未详尽描述的技术内容均为公知技术。

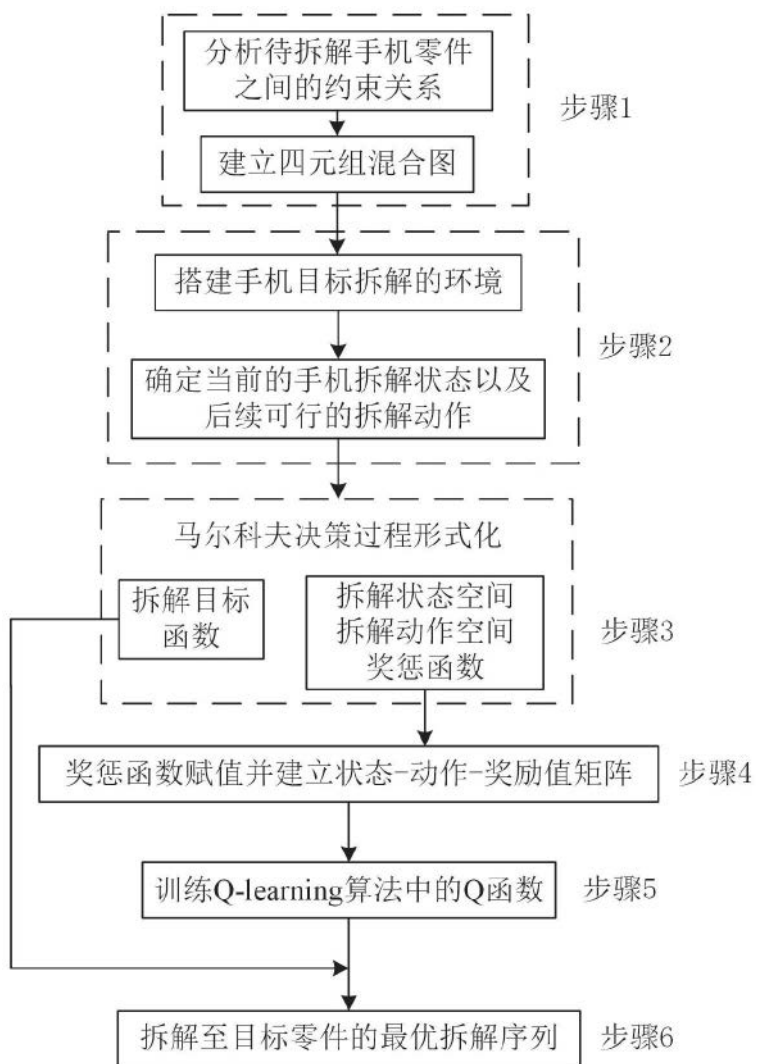


图1

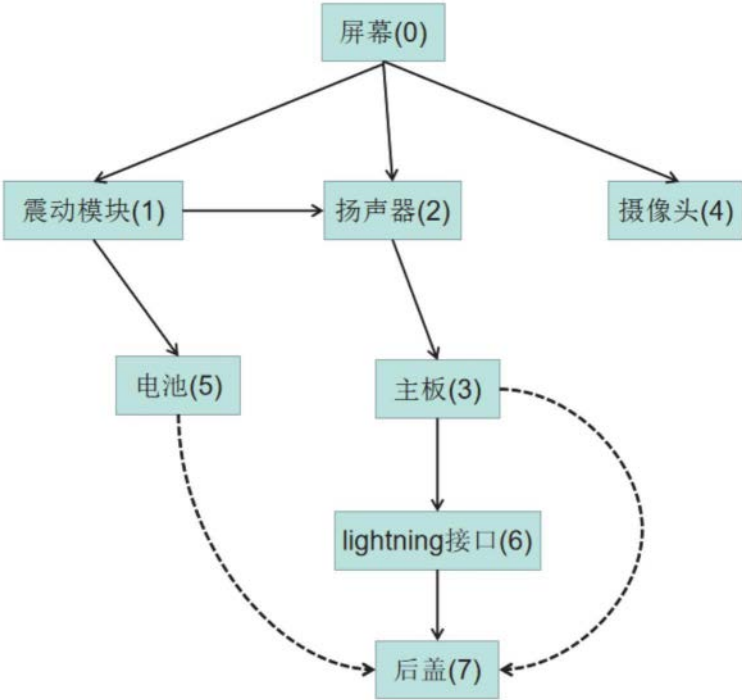
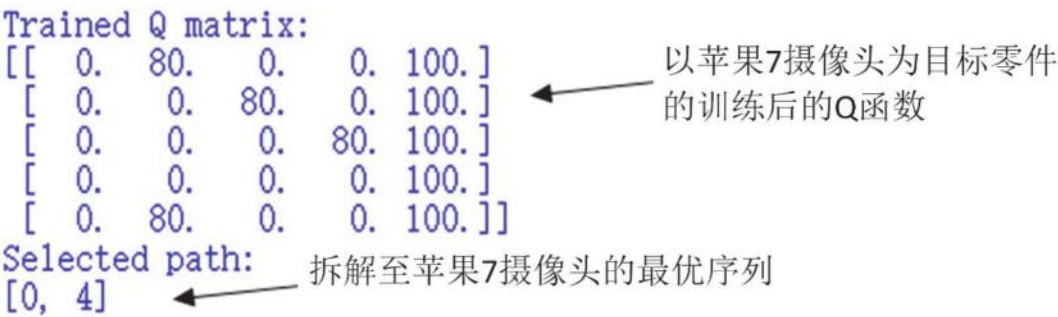


图2



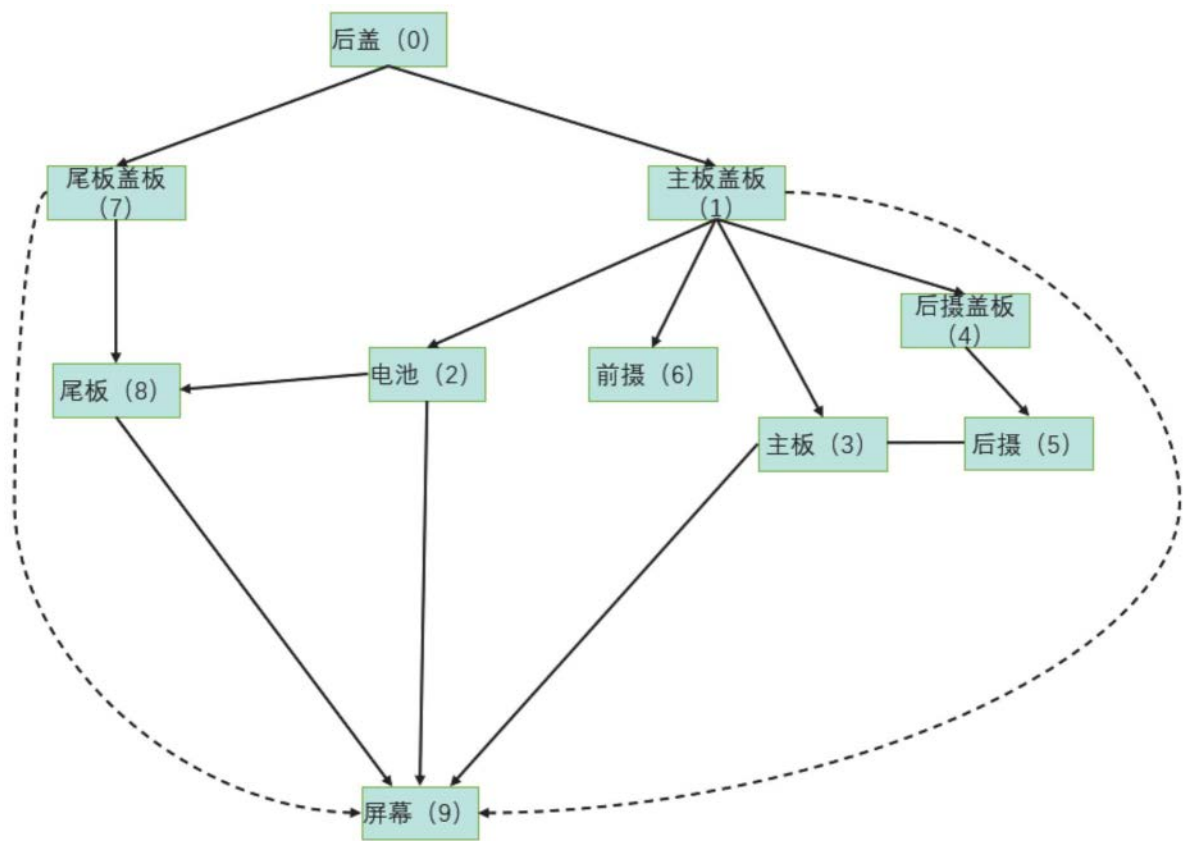


图5

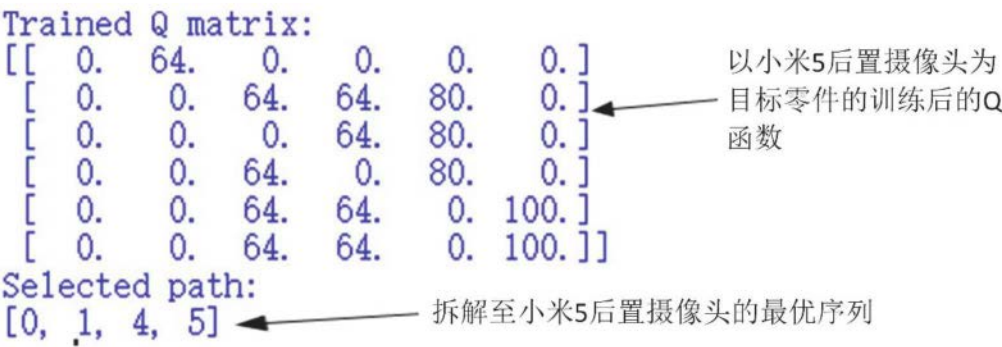


图6

Trained Q matrix:

[	0.	80.	0.	0.	0.	0.]
[	0.	0.	80.	100.	80.	0.]
[	0.	0.	0.	100.	80.	0.]
[	0.	0.	80.	100.	80.	0.]
[	0.	0.	80.	100.	0.	80.]
[	0.	0.	80.	100.	0.	0.]

Selected path:

[0, 1, 3]

以小米5主板为目标零件的训练后的Q函数

拆解至小米5主板的最优序列

图7