

基于分级代理的智能家庭网络模型研究^①

么 刚^{②*} 张 武^{③**} 王劲林^{**}

(* 中国科学院研究生院 北京 100049)

(** 中国科学院声学研究所 北京 100190)

摘 要 针对现有家庭网络中智能设备不能有效学习家庭用户习惯,致使不能满足用户的个性化服务质量的问题,给出了一种基于分级代理的智能家庭网络模型。首先给出了一种智能家庭网络设备的形式化描述,以此为基础提出了分级代理的智能家庭网络模型:全局 Agent 通过对家庭中的长期数据进行学习,总结出一定的服务规则,指导设备 Agent 根据家庭成员的生活习惯改变工作方式;设备 Agent 利用强化学习算法,自主学习,解读家庭环境的状态变化,并做出最优选择。该模型的应用实例及仿真结果表明,应用此模型,设备可以学习用户习惯,为用户提供个性化服务。

关键词 智能家庭网络,强化学习,多级代理

0 引 言

随着计算机技术、通信技术和网络技术的发展,以及消费水平的不断提高,人们更加注重生活质量,对家居环境的舒适性、便利性、安全性、高效性等提出了更高层次的要求,智能家居随之出现。目前的智能家居系统主要包括照明控制系统、电器控制系统、安防门禁系统、报警系统和远程控制系统等。上述系统利用 3C(计算机、通讯、消费电子产品)技术,使用户可以随时随地获得透明的数字化服务。然而在现有的事件触发模式下,设备各自完成相对独立单一的任务,设备之间缺乏有效的协调合作,难以完成更为复杂的任务。为此 Tsai 等人首先提出了 MASSIH^[1],这是一种智能家庭网络的多代理架构,之后文献[2,3]中又分别提出了 MAIH^N 架构和 MAIH^S 架构,但是上述架构在完成复杂任务时,仍然需要用户利用用户接口(UI)进行全程干预。

为最大程度地实现设备的智能化,尽可能地减少用户对智能家庭网络的人为干预,本文提出了一种智能家庭网络的分级代理模型,通过全局代理(Agent)和设备 Agent 的分工合作,实现家庭网络的智能化。全局 Agent 负责对家庭中的长期数据进行学习,总结出一定的服务规则并不断修正,指导设备

Agent 根据家庭成员的生活习惯改变工作方式;设备 Agent 利用强化学习算法,自主学习,解读家庭环境的状态变化,自动做出调整并做出最优选择。本文根据现有的家庭网络协议,抽象出一种统一的家庭网络的形式化描述,并以此为基础展开了研究,同时分析了智能设备的需求,选择了适合的机器学习算法,并对其作了介绍,最后给出了应用实例和仿真结果。

1 家庭网络的形式化描述

1.1 现有家庭网络标准

随着家庭网络研究的日益深入,各大标准协会以及通信、家电厂商纷纷从不同的角度推出了家庭网络标准。从设备的互通性角度出发制定标准的组织有数字生活联盟(DLNA)、闪联(IGRS)、e 家佳(ITopHome)、泛在开放平台论坛(UOPF);从中间件角度出发进行研究的组织和技术有通用即插即用(UPnP)、开放服务网联盟(OSGi)、HAVi、Jini。其中对我国智能家庭网络有较大影响的是由 Microsoft 首先提出的 UPnP^[4]以及由联想牵头提出的现已成为 ISO 国际标准的 IGRS^[5]。

UPnP 旨在“为来自不同厂商的消费电子、智能家电和移动设备提供简单和高效的连接”。UPnP 协

^① 863 计划(2007AA01Z238)资助项目。

^② 男,1981 年生,博士生;研究方向:家庭网络;E-mail: yaog@dsp.ac.cn

^③ 通讯作者, E-mail: zhangw@dsp.ac.cn

(收稿日期:2008-09-16)

议中设备的基本行为包括设备寻址、设备发布、设备发现、设备控制、事件订阅和用户接口,通过基本行为可以使设备在网络中发布自己可提供的服务,搜索自己所需的服务,并对自己获得的服务进行控制。

IGRS 是“信息设备资源共享协同服务标准”的简称,是“新一代网络信息设备的交换技术和接口规范”,在通信及内容安全机制的保证下,“支持各种 3C 设备智能互联、资源共享和协同服务”。IGRS 的基础协议包括以下方面:资源发现与发布机制、设备配置管理机制、服务访问控制机制、数据获取分发机制、安全规范。在此基础协议之下,IGRS 设备可以解决互联互通、资源共享和协同合作的问题。

1.2 智能家庭网络的形式化描述

在家庭网络中,众多的智能设备使用各自的家庭网络标准和协议,成为了泛在的智能家庭网络研究的最大障碍之一。文献[6]以可扩展标记语言(XML)和 BACnet 协议为基础给出了一种家庭网络的统一描述,但并没有兼顾已有家庭网络协议;文献[7]将家庭网络定义为由人、设备和管控规则构成的信息系统,是从信息系统的角度给出了家庭网络的形式化描述。本文从智能设备互联互通及互操作的角度出发,分析了 UPnP、IGRS 等协议,总结出各个协议在基础结构和设备行为上的共性,抽象出了一种统一的家庭网络的形式化描述,以此为基础进行分级代理模型的研究。

(1) 家庭网络的组成

从家庭网络设备之间的互连互通和协同合作角度,家庭网络可以被划分为主控设备、被控设备和描述。描述包括设备描述和服务描述,设备描述包括设备可提供的服务,设备属性和图表等,服务描述包括服务的动作、状态变量、输入输出参数。由主控设备调用被控设备的服务,完成所需功能。主控设备和被控设备不是绝对的,在不同的场景下主控设备和被控设备可以相互转化。

(2) 设备的行为

分析 UPnP、IGRS 等家庭网络协议可以发现,在家庭网络中,实现设备的互操作性,设备需要具备设备寻址、设备发布、设备发现、设备控制、事件订阅等基本功能。设备寻址,使设备获得能够在家庭网络中使用的地址;设备发布,将设备自身的设备信息和所能提供的服务信息发布给其他智能设备;设备发现,使设备能够在网络中发现自己感兴趣的设备;设备控制,使主控设备能够控制被控设备完成所需功能;事件通知,使设备能够及时获知其他设备的状态

变化。

(3) 设备的形式化描述

智能设备的形式化描述如下:

$$S = \langle O, A, V, F, D \rangle$$

O 表示设备对象有限集合, $O = \{o_1, o_2, o_3, \dots, o_n\}$ 。每一个设备或设备中的功能模块都可以看作是一个对象。

A 表示对象属性集合 $A = \{a_1, a_2, a_3, \dots, a_m\}$ 。一般对象的属性集合可以分为不相交的 3 个子集: A_{in} (条件属性)、 A_{out} (目标属性) 和 A_{con} (固有属性)。

V 表示对象属性的定义域, $V = \bigcup_{a \in A} V_a$ 。

F 表示对象所具有的服务, $F = \{f_1, f_2, f_3, \dots, f_q\}$ 。服务利用属性集合中的条件属性得到目标属性。

D 表示设备联动的领域知识集合, $D = \{d_1, d_2, d_3, \dots, d_p\}$ 。

尽管表述不尽相同,但是 O 、 A 、 V 、 F 都是现有的以互操作为目标的家庭网络协议所共有的,而 D 是为实现智能化,设备经过长期学习得到的相关领域的知识。

(4) 服务事件的形式化描述

在设备对家庭网络用户生活习惯的学习过程中,需要对家庭网络服务事件进行分析,得到家庭用户的长期生活数据,进而指导设备调整其工作,使之符合用户习惯。我们以七元组 $\langle S, O, a, v_a, f, t_s, t_e \rangle$ 来描述家庭网络服务事件。其中: S 是服务所属设备; O 是提供服务的对象; a 是与服务相关的属性,这里的 a 是指发生变化的属性,可以是输入属性 a_{in} 或者输出属性 a_{out} ; v_a 是与 a 对应的属性的值; f 指的是具体发生的服务事件; t_s 指的是服务发生的起始时间, t_e 指的是服务事件结束的时间。当服务状态发生变化时,将反映在 a 及 v_a 上。

(5) 服务联动性的定义

在服务事件的基础上,进一步定义服务事件的联动性。在任务模式下,设备 Agent 之间的联动关系可以分为冲突和合作。冲突是指设备在提供各自的服务 f 时属性 a 相冲突,这里的 a 既可能是输入属性 a_{in} 也可能是输出属性 a_{out} ,我们判断设备服务相冲突的依据是服务的作用时间 $[t_s, t_e]$ 相交叠,并且服务事件的相关属性 a 相冲突。

合作又分为横向合作和纵向合作。横向合作是在同一时间段之内,由多个设备协作,完成一个任

务,我们判断一组设备服务横向合作的依据是 $[t_s, t_e]$ 相交叠,并且 t_s 是在短时间内连续发生;纵向合作是将任务划分为多个子任务,由不同的设备依次完成,我们判断一组设备纵向合作的依据是 $[t_s, t_e]$ 依次发生,并且前一个事件的 t_e 和后一个事件的 t_s 在短时间内发生。

2 智能设备学习算法

在家庭网络中,为完成特定任务,设备在协同工作时可能构成冲突或合作的关系,此时家庭成员需要根据任务特征以及家庭网络状态对设备的服务进行调整。然而在智能家庭网络中,希望智能设备能够通过自主学习,解读外界环境的状态变化,自动做出调整并做出最优选择。从而在减少家庭成员参与的情况下,自动高效地完成家庭网络任务。

强化学习(RL)^[8]是作为自治系统中的一种灵活、自适应的控制方法,在机器学习领域得到了广泛的应用。强化学习包括策略、奖励函数、值映射、环境空间和动作空间 5 个基本要素,其中策略是由环境空间到动作空间的映射规则;奖励函数是动作作用于环境之后,环境给智能主体的反馈;值映射是环境空间到动作空间映射的量化表示。强化学习的目标就是智能主体通过与环境不断交互获得知识,以逐步调整形成最佳的行为策略。相比其他的机器学习方法,强化学习以其自学习和在线学习的特点成为研究的热点,得到了广泛的关注和发展,并且越来越多地与多 Agent 技术相结合以解决实际问题^[9-11]。

受此启发,本文将 RL 中应用最为广泛的 Q 学习^[12]算法引入设备 Agent,通过设备和环境的交互,逐渐获得最佳的行为策略,从而使家庭网络用户自动地获得最佳的家庭网络服务体验。

强化学习是要通过不断地“试错”并迭代更新映射值表的方式,使智能主体获得一个最优策略 π^* : $S \rightarrow A$, 使得智能主体选择的动作能够获得最大的长期积累回报,其中 S 为状态集, A 为动作集。具体实现是一个迭代过程,每一轮的迭代中,学习者感知环境状态,并根据当前策略 π 选择动作 $a \in A$ 作用于环境,环境由此变化为 $s' \in S$, 同时产生 t 时刻的奖励值 $r_t = r(s, a)$ 反馈给学习者;学习者根据 s' 和 r_t 更新其策略 π , 进入下一轮迭代。

长期积累回报可表示为

$$V_{\pi}(s) = E\{r_t + \gamma \times r_{t+1} + \gamma^2 \times r_{t+2} + \cdots | s_t = s, \pi\} \quad (1)$$

其中 $\gamma \in (0, 1)$ 为折扣系数,根据文献[13]及 Bellman 等式^[14],式(1)的最大值为

$$V^*(s) = \max_{s' \in S} \sum p(s' | s, a) [r(s, a) + \gamma V^*(s')] \quad (2)$$

Q 学习可以在 $p(s' | s, a)$ 和 $r(s, a)$ 未知的情况下找到最优策略 π^* , 应用最为广泛。Q 学习通过以下迭代规则得到:

$$Q_{k+1}(s, a) = (1 - \alpha_k) Q_k(s, a) + \alpha_k [r + \gamma \max_{a'} Q_k(s', a')] \quad (3)$$

其中 k 为迭代次数, α 为学习率, γ 为折扣率, Agent 由状态 s 经过动作 a 后到达状态 s' 所接受的瞬时奖赏值。

3 智能家庭网络的分级代理模型

为实现设备的学习性,令每一个设备对应于一个 Agent。然而家庭网络中,设备众多,每个设备状态的改变和设备动作的发生都会使整个家庭网络的状态发生变化。要使每一个设备的 Agent 维护整个家庭网络的状态,获得家庭网络全局性的知识是不必要的,也是不可能的。为此提出了一种智能家庭网络的分级代理模型,如图 1 所示。全局代理解决的问题是家庭网络的子服务该由哪个设备提供以及什么时候提供,而设备代理解决的问题是在设备联动的层面如何更好地提供服务。全局 Agent 需要维护整个家庭网络的状态,进行仲裁知识和协作知识的学习,而设备 Agent 需要学习与本设备相关的联动知识。

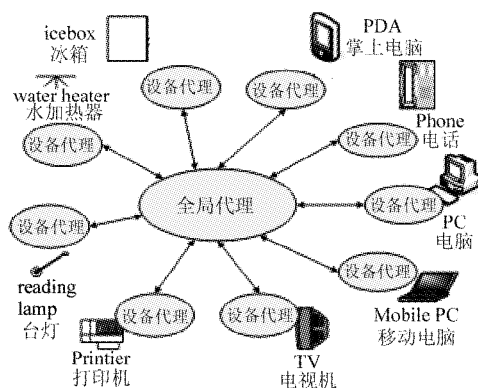


图 1 基于分级代理模型的家庭网络示意图

3.1 全局代理

(1) 全局代理功能模块

全局代理是一个特殊的虚拟设备代理,具备设

备的基本功能:设备寻址、设备发布、设备发现、事件订阅和设备控制。另外全局代理也可以接受用户的任务设置,并向用户报告任务的执行情况。在全局代理的内部,分为通信模块、学习模块、决策模块、状态监控模块,并进行事件记录。全局代理模块示意图见图 2。

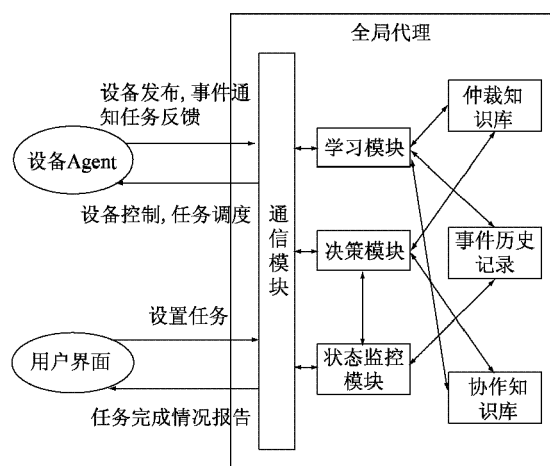


图 2 全局代理模块示意图

通信模块用以连接全局代理和外界环境。首先通信模块要负责与家庭网络中的服务器以及设备代理之间的通信,完成全局代理的设备寻址、设备控制、时间通知等动作。另外通信模块还负责和用户界面进行交互,负责执行由用户下达的任务,并向用户报告任务的完成情况。

状态监控模块从通信模块中分析得到网络中其他设备的状态,一方面为决策模块提供决策依据,另一方面将家庭网络服务事件记入事件历史记录,作为用户生活习惯的长期数据,为学习模块提供素材。状态监控模块通过监控其他设备的设备发布和事件通知维护整个家庭网络的状态,并以 $\langle S, O, f, a, v_a, t_s \rangle$ 维护当前活跃的动作。当服务结束时,记入事件记录。

决策模块是全局代理的核心,由决策模块来控制其他设备的活动,促成设备间的合作,解决设备间的冲突。决策模块根据家庭网络当前状态以及仲裁知识和协作知识进行决策。当决策模块探测到新服务 f_{new} 和现有服务 f_{old} 相冲突时,决策模块以事件通知的形式告知 f_{new} 和 f_{old} 的所属设备可能因属性 a 与对方相冲突,具体的冲突解决方案由设备代理经由强化学习做出。服务的合作由全局 Agent 控制,全局 Agent 负责维护家庭网络的全局状态,若合作

的触发条件发生,则对相关的设备 Agent 进行设备调用,维护任务的完成情况,并向用户报告。

事件历史记录以服务事件的形式记录家庭网络发生的服务事件,以 t_s 作为索引对发生的服务事件进行排序。事件的历史记录由状态监控模块提供数据,并作为学习模块的学习素材。

学习模块以事件历史记录作为原始素材,根据对设备冲突和设备合作的定义,判断设备冲突和设备合作,分析家庭用户对设备冲突和设备合作的解决方法,并作为仲裁知识和协作知识存入仲裁知识库和协作知识库。

(2) 全局代理学习算法

学习模块根据服务联动性定义判断服务是否冲突。例如:电视的“音量”为 TV 设备的输出属性 a_{out} , 此时有 IP 电话接入,“音量”同时也是 IP Phone 设备服务的输出属性 a_{out} , 此时学习模块初步判定电视和电话的服务是冲突的。如果用户发起了调低电视音量至 v_a 的动作,那么学习模块将此动作作为解决 TV 和 IP Phone 冲突的解决方案之一。经过对家庭用户长期生活数据的学习就可以得到 TV 和 IP Phone 冲突的最佳解决方案。另外,由于学习模块对属性冲突的判断有赖于对属的语义描述,因此学习模块还需要处理由于语义描述的不同而造成的错判和漏判。如果服务 A 和服务 B 并不满足服务冲突的定义,但在服务 A 和服务 B 同时开启时,用户总是会对 A 或 B 进行调整,则学习模块认定 A 和 B 相冲突,并自动学习用户对该方案的处理方式。反之,如果服务 A 和服务 B 的服务符合服务冲突的定义,但是用户并没有对服务 A 和服务 B 进行调整操作,则学习模块认为服务 A 和服务 B 不相冲突。

设备服务的合作分为横向合作和纵向合作。学习模块对于合作的判断也依据服务的联动性定义。在横向合作中,学习模块还需要找出一个合作发起服务,如“打开高清电影播放器→打开低音炮→调暗光线”是一组横向合作,其中“打开高清电影播放器”就是合作发起服务,可以由它发起“打开低音炮”和“调暗光线”的服务,但是不能由“调暗光线”发起“打开高清电影播放器”的服务。纵向合作依赖于用户的指导。

3.2 设备代理

(1) 设备代理功能模块

设备代理首先具备智能设备的基本功能,还可以接受全局 Agent 的任务调度,并进行任务反馈。在设备代理内部有行为规则,用以指导设备服务、设

(4) 全局代理将 FTP 下载服务记入状态监控模块。

(5) IPSTB 机顶盒开机,进行设备发布,全局代理监测到设备发布消息。

(6) 全局代理向 IPSTB 设备进行事件订阅,以获得设备的状态变化。

(7) IPSTB 开启流式高清播放服务,向全局代理进行事件通知。

(8) 全局代理将 IPSTB 流式高清播放服务记入状态监控,发现 PC 的 FTP 下载服务和 IPSTB 的流式高清播放服务的“ α_{in} = 下行带宽”相冲突。

(9) 全局代理检查仲裁知识中的“过滤规则”,发现冲突确实存在。

(10) 通知 PC 和 IPSTB 代理,“ α_{in} = 下行带宽”相冲突。

(11) PC 代理和 IPSTB 代理之间以事件通知的形式,进行协商,PC 代理和 IPSTB 代理进行学习以获得家庭网络的最优服务质量。

4.2 仿真结果

当 PC 及 STB 收到“ α_{in} = 下行带宽”冲突消息时,按照强化学习进行冲突协商。表 1 所示为仿真参数。

表 1 仿真参数

总下行 带宽	学习 率	折扣 率	迭代 速率	动作空 间大小	环境空 间大小	服务 优先级
40Mbps	$\alpha=0.5$	$\gamma=0.8$	0.1s/次	$N=10$	$S=10$	0.5:0.5
40Mbps	$\alpha=0.5$	$\gamma=0.8$	0.1s/次	$N=10$	$S=10$	0.1:0.9

服务的满意度由设备按照服务特性估算得到,TFP 的服务满意度由下载速度衡量;标准清晰度电视(SDTV)的服务满意度由解码器的充盈度衡量。在实际环境中,服务满意度和相关输入属性的关系由设备实时计算得到。为方便仿真,图 5 给出了理想环境下 TFP 和 SDTV 的下行带宽/服务质量的评估曲线。

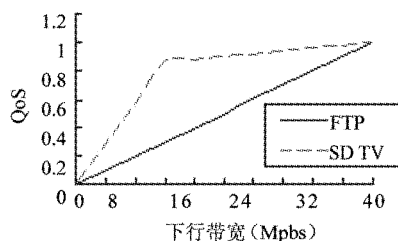


图 5 FTP/SD TV 的服务质量评估曲线

图 6 为经强化学习,FTP 服务占用下行带宽曲线,图 7 为 FTP/SDTV 的整体满意度曲线。

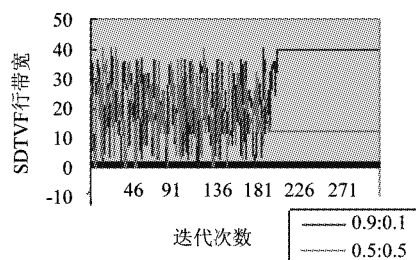


图 6 SDTV 下行带宽曲线

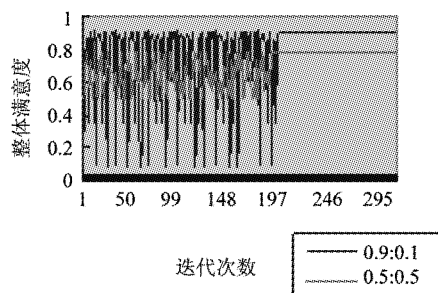


图 7 FTP/SDTV 的联动整体满意度曲线

仿真结果分析:

(1) 收敛速度:Q 学习的收敛速度取决于学习率、折扣率等参数,Q 学习算法本身体现了收敛速度的参数是收敛迭代次数,本仿真参数下,收敛发生在 200 次迭代左右。具体到家庭网络环境,学习时间取决于收敛迭代次数以及迭代速率。假设设备可以完成 10 次/s 以上的迭代计算,则学习时间可以控制在 20s 之内。并且在进行一次 Q 学习之后,智能设备会记录下相关的联动知识,在遇到相同冲突时根据联动知识解决,不再进行 Q 学习。因此本 Q 学习的收敛速率基本可以满足家庭网络的要求。

(2) 个性化:本仿真给出了优先级为 0.5:0.5 和 0.9:0.1 两组数据,得到了不同的结果。当优先级为 0.5:0.5 时,SDTV 所占带宽为 12Mbps 时,获得最大整体满意度。当优先级为 0.9:0.1 时,SDTV 所占带宽为 40Mbps,获得最大整体满意度。说明在本模型下,根据不同用户的习惯,设备可以为用户提供不同的用户体验。

5 结论

本文在分析主要家庭网络协议的基础上,给出了智能家庭网络的实行化描述,以此为基础提出了

家庭网络的分级代理模型,并在设备代理中引入了强化学习算法,通过家庭中的长期数据进行学习,使家庭设备根据家庭成员的生活习惯改变工作方式,实现家庭网络的智能化。

参考文献

- [1] Tsai C F, Wu H C. MASSIHN: a multi-agent architecture for intelligent home service. *IEEE Transactions on Consumer Electronics*, 2002, 48(3): 505-514
- [2] 韩江洪,江波,度诗研等. 一种基于多 Agent 系统的智能家庭网络研究. *电子科技大学学报*, 2005, 2(34): 233-235, 276
- [3] 李效华,赵现钢. 一种基于多 Agent 的智能家居系统研究与设计. *计算机工程*, 2006, 21(32): 233-234, 257
- [4] UPnP Forum. UPnP™ device architecture 1.0 version. <http://www.upnp.org>: UPnP Forum, 2003
- [5] 闪联标准工作组. 信息设备智能互联与资源共享协议 1.0 version. 北京: 闪联标准工程作, 2003
- [6] 王中明. 基于 XML/BACnet 的家庭网络系统研究: [博士学位论文]. 武汉: 华中科技大学, 2006
- [7] 蒋建文. 基于多 Agent 的信息家居系统研究: [博士学位论文]. 合肥: 合肥工业大学, 2003
- [8] 高阳, 陈世福, 陆鑫. 强化学习研究综述. *自动化学报*, 2004, 30(1): 86-100
- [9] 何佳, 程显毅, 李晓薇. 基于用户反馈的智能合作过滤模型的研究. *智能系统学报*, 2007, 2(1): 59-63
- [10] 陈阳舟, 张辉, 杨玉珍等. 基于 Q 学习的 Agent 在单路口交通控制中的应用. *公路交通科技*, 2007, 24(5): 117-120
- [11] 沈晶, 顾国昌, 刘海波. 基于多智能体的 Option 自动生成算法. *智能系统学报*, 2006, 1(1): 84-87
- [12] Watkins C J C H, Dayan P. Q-learning. *Machine Learning*, 1992, 8(3): 279-292
- [13] Watkins C J C H. Learning from delayed rewards: [Ph. D dissertation]. Cambridge: Cambridge University, 1989
- [14] Bellman R E, Dreyfus S E. *Applied Dynamic Programming*. New Jersey: Princeton University Press, 1962

Research on intelligent home network model based on multi-level agents

Yao Gang^{* **}, Zhang Wu^{**}, Wang Jinlin^{**}

(^{*} Graduate University of Chinese Academy of Sciences, Beijing 100049)

(^{**} Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190)

Abstract

This paper presents an intelligent home network model based on multi-level agents to solve the problem that the existing intelligent devices in home networks can not study users' habits efficiently so that they can not satisfy users' personalized QoS. A formal description for devices in intelligent home networks is given before the model is established. The overall agent in an intelligent home network summarizes the rules of service through the study of the long-term data of the family, which guide the work of equipment agents according to the life habits of the family. The equipment agents learn and detect the change of the family environments automatically and make the optimal choices using the reinforcement learning algorithms. An application scene of the multi-level agent model and the simulation result show that, by applying this model, the devices can study the habits of the user and provide personalized service to the user.

Key words: intelligent home network, reinforcement learning, multi-level agents