Shutter，机器人摄影师：利用行为树实现真实场景中的公共人机交互

Alexander Lew计算机科学 耶鲁大学康涅狄格州纽黑文
a.lew@yale.edu

Sydney Thompson计算机科学 耶鲁大学康涅狄格州纽黑文
sydney.thompson@yale.edu

Nathan Tsoi计算机科学 耶鲁大学康涅狄格州纽黑文
nathan.tsoi@yale.edu

Marynel V´azquez计算机科学 耶鲁大学康涅狄格州纽黑文
marynel.vazquez@yale.edu

arXiv:2302.00191v1 [cs.RO] 1 Feb 2023

摘要——在真实场景中部署交互系统，需要系统能够适应实验室环境中未曾遇到的情形。本文介绍了我们在部署公共交互系统的过程中，对架构选择如何影响行为的可复用性和反应性的经验。具体而言，我们介绍了 Shutter：一位机器人摄影师，也是一个用于公共交互的平台。在设计 Shutter 的架构时，我们着重考虑了其在真实场景部署中的适应性，同时开发了一个可复用的平台，以促进未来的公共人机交互研究。我们发现，行为树能够提升反应性，尤其是在群体场景中，并有助于推动可复用行为的设计。

索引词—**人机交互、人机交互**

一、引言

面向公众的交互智能体面临着受控环境中不会出现的独特挑战。在真实场景中部署需要具备鲁棒性，既要适应环境，也要适应用户，因为公共互动经常出现意料之外的情况，要求智能体灵活应对，而且身边可能没有专家纠正错误。应对这些多样情况的一种常见方法，是为这些情况定义特定的机器人响应 [1]。

公共空间的复杂性推动了更完善的感知与交互方法的发展。例如，群体检测 [2]是提升机器人在社交环境中活动能力的一项重要能力。随着这些智能体的模态数量增加、能力增强，以简单而鲁棒的方式协调机器人活动，是设计引人入胜的交互的关键一步。机器人的底层决策架构塑造了其交互动作的设计，尤其关系到这些动作的模块化与可复用性。

在这项工作中，我们介绍了 Shutter，一位机器人摄影师，并分享了我们在实现行为架构方面的经验，以便与周围环境进行响应迅速且可复用的交互。本工作的主要贡献是首次提出

本研究部分由美国国家科学基金会（NSF）资助，资助编号为（IIS-1924802）。作者感谢众多为 Shutter 作出贡献的学生，包括 B. Lyng-Olsen、T. Adamson、V. Del Carpio、Y. Milkessa、E. Gorevoy、A. Narcomey、Q. Zhang 和 S.Gillet。



图 1：Shutter，一款为公共互动而设计的机器人摄影师平台。

将 Shutter 作为公共交互平台，并讨论了在设计适用于真实环境部署的模块化响应式系统时所获得的经验教训。

II. 相关工作

针对部署于公共环境中的社交机器人，人们提出了多种控制架构，包括认知架构、有限状态机和行为树。认知架构既对交互响应进行建模，也对其进行选择[3], [4]。有限状态机是一种规定机器人活动的经典方法，已应用于各种领域，包括引导博物馆导览 [5] 和搜救 [6]。有限状态机架构由定义动作的状态，以及表示对观测或用户输入所作反应的状态转换组成。状态转换条件定义了从每个状态出发所采取的状态转换。示例见图 2a。状态之间的转换可以由用户通过按钮、触摸屏界面 [7],或触觉响应 [8],输入

用户的语言响应 [1],或机器人成功完成某个动作 [9]来触发。

状态机在公共交互机器人中有许多应用实例。Directions Robot [1] 采用了基于有限状态的对话管理器，根据场景分析控制对话动作，以便为用户提供导航指引。在 Robovie[8], 于购物中心的部署中，机器人活动由控制模块根据上下文信息和预先设定的情节规则进行协调。这些研究也探讨了在公共环境中部署机器人所面临的挑战，例如如何应对用户出乎意料的动作，以及如何观察用户自发参与和退出互动。在社交机器人学领域，Bauer 等人 [7] 实现了一个有限状态机，用于定义向用户询问前往目标位置的路线的交互。 [9] 也采用有限状态机来协调一台担任导游的移动机器人的高级活动。

此前，行为树也曾用于设计公共交互。行为树是另一种用于指定自主智能体可执行动作的结构 [10], [11]。叶节点表示智能体可执行的行为。生成语音、收集传感器数据以及更新内部数据结构，都是行为的实例。行为的执行顺序由内部节点（即控制流节点）决定。后备、并行和顺序是控制流节点的主要类型。行为树会从根节点开始，以固定的时间间隔进行遍历，这些间隔称为 tick 事件。每次收到 tick 时，行为都会返回成功、运行中或失败状态，父控制流节点则利用该状态确定下一个接收 tick 事件的节点。图2b展示了行为树中三种主要内部节点的示例，该行为树定义了与 Shutter 进行基本拍照互动的流程。

在 [12]，中，我们将行为树作为一种框架进行评估，用于编写真实场景中的儿童与机器人互动脚本。这项工作与我们提出的方法有一个主要区别：机器人的群体感知程度不同，因为 [12] 中的反应式行为假设用户始终在场。

III. SHUTTER，机器人摄影师

我们的社交机器人 Shutter 是一位机器人摄影师，作为面向公共互动的平台而开发；我们希望将其用于走廊、门厅或展厅等场景，以开展多种拍照互动。为此，我们注重 Shutter 架构的可复用性。机器人的主要用途虽然是拍照，但我们设计 Shutter 时也让它能够支持多种互动。在首个版本中，Shutter 会检测潜在的互动对象、拍摄摆拍照片，并向用户展示照片。

A. 硬件

图 1 展示了我们的机器人平台。Shutter 安装在一辆推车上，可部署到各种室内地点。该机器人是一台 Trossen Robotics WidowX 机械臂，并加装了面部显示屏。机器人头部还安装了一台 Intel RealSense RGB-D相机，可提供自我中心感知。推车上安装的 Azure Kinect 传感器可从第三人称视角观察环境，并进行人体追踪。Kinect 安装在大型显示屏上方，为用户提供辅助视觉界面。机器人下方的扬声器可播放音频，包括生成的语音。用户可以按下推车正面的三个发光按钮中的任意一个，向 Shutter 提供反馈，例如从菜单中选择选项。

B. 软件

我们创建了定制软件，用于控制底层动作，例如运动和面部表情渲染，并实现用于指定交互的高层控制架构。Shutter 还利用现有库，例如提取 AzureKinect 的人体追踪数据。为促进更多面向公众的交互系统的开发，Shutter 的控制架构和动作定义源代码已公开。

群体感知是为 Shutter 创建自适应响应的关键。群体检测算法 [2] 会将人群划分为多个社交群体，其中包括 Shutter 所在的社交群体。在实践中，检测群体能够实现交互的动态变化。例如，Shutter 目前会检测互动对象的数量，并根据群体规模调整问候语。如果有 3 个人与 Shutter 互动，如图 1 所示，Shutter 会向这群人问道：“你们三位想让我为你们拍张照片吗？”。与对群体本身进行建模的情况一样 [2], , 社交群体中的人数事先并不可知，因此 Shutter 更复杂的群体感知动作需要适应群体规模的变化。此外，群体规模也可能在交互过程中发生变化。因此，在评估 Shutter 的软件架构时，我们优先考虑其适应群体规模和当前互动对象变化的灵活性。

IV. 架构比较

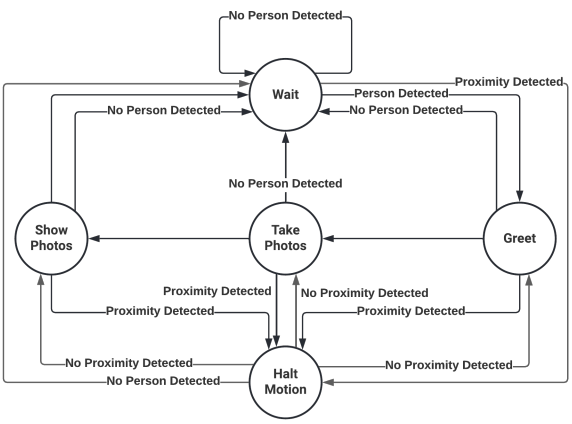
在为实现预期交互而定义 Shutter 的软件架构时，我们考虑了两种方案：有限状态机和行为树。尽管二者在实践中具备相似的能力，但各自架构所规定的结构会影响动作的复用方式，以及实现反应式行为的难易程度。

A. Shutter 基本交互

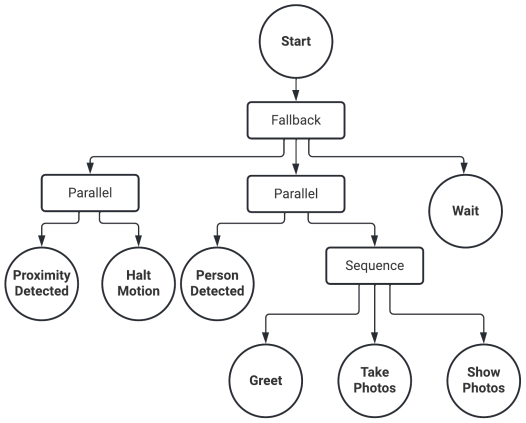
当 Shutter 的 Kinect 传感器检测到附近至少有一个人时，交互便会开始。机器人会自我介绍，并询问：“您想让我为您拍照吗？”如果有 $n > 1$ 位用户，则会问：“您想让我为你们拍照吗？”在获得用户同意后，Shutter 会告知互动者即将为他们拍照，并使用机载 RealSense 相机拍摄三张照片。照片会显示在屏幕上，同时 Shutter 会称赞照片，例如说道：“您在这张照片里看起来真棒。”

B. 有限状态机架构

最初，我们使用有限状态机实现了 Shutter 的交互。在我们的有限状态机中，节点表示



(a) 表示 Shutter 基本交互的有限状态机。



(b) 表示 Shutter 基本交互的行为树。

图 2：使用有限状态机和行为树实现的 Shutter 行为架构。

Shutter 执行的高级活动，状态转换则表示需要改变机器人活动的人类响应。图 2a 展示了有限状态机。沿着状态之间的状态转换，可以从语义层面理解 Shutter 的活动。有限状态机的一个主要优点是，这些明确的状态转换使 Shutter 的响应过程更加透明。

然而，随着有限状态机中的状态数量不断增加，我们发现，理解状态转换以及针对变化的环境创建反应式响应变得困难得多。这是因为我们通过显式状态转换来定义行为交互。因此，像“停止运动”这样的状态必须能够从交互中的每个位置到达，这意味着需要有多条状态转换指向该状态。这增加了有限状态机所需的状态转换数量。总的来说，我们发现，这种额外负担限制了初始有限状态机中子图的可复用性。在尝试扩展有限状态机以适应不同的交互设计时，我们发现很难高效地修改默认交互。

C. 行为树架构

为了解决我们对有限状态机的顾虑，我们为 Shutter 实现了一种行为树，其中的行为要么推进正在进行的交互，要么更新交互上下文。协调这两类行为，使我们能够实现反应式机器人控制。例如，检测到附近的人以启动基本交互，属于更新交互上下文的行为。图 2b 中的“检测到人员”节点就是这种更新的一个示例。

开发行为树时，一个关键考量是行为可复用性。例如，用于启动交互的同一近距离检测行为，在检测不到用户时也会被重复用于停止交互。定义更复杂行为的子树（例如记录按钮按压）也可以重复使用。这使我们能够通过新的方式连接现有子树，更轻松地定义新的交互。下一节将介绍一些具体的交互场景在这些场景中，相较于有限状态机，我们发现行为树更有优势。

D. 使用行为树的示例

1）交互中途放弃：与实验室环境不同，面向公众的交互系统必须能够妥善应对用户中途放弃交互的情况，因为用户可能会分心，或在交互过程中需要离开。对于 Shutter 而言，这种情况表现为交互进行期间，推车上的 Kinect 未检测到人员，系统随后返回等待状态。

使用有限状态机处理交互中途放弃需要采用以下两种方案之一：1）用户离开后，返回起始状态；或 2）在用户不在场的情况下完成交互。如图 2a 所示，为处理中途放弃而添加状态转换，需要从每个状态都添加一条返回等待状态的状态转换。另一种方案是为每个状态设置超时选项，使有限状态机返回等待状态。然而，这种方案要求交互结束前的每个状态都等待至超时。虽然这些方案易于理解，但都需要添加大量状态转换，或设置较长的重置周期。

与有限状态机不同，行为树旨在从此类失败情形中快速恢复。如图2b所示，如果“检测到人员”节点失败，父级并行块也会失败，而前面的回退节点则默认进入等待状态。使用行为树处理交互中途放弃，只需进行一项架构更改，即添加并行块；而修改有限状态机则需要为每个节点添加额外的状态转换或超时。

2）更高优先级的中断：决策架构必须可靠处理的另一种真实场景是高优先级中断或后备。在自然交互过程中，某种情况可能要求机器人立即采取特定行动。例如，安全回退要求机器人中断计划中的活动，以降低风险。另一个例子是网络连接丢失时暂停交互。一旦优先级条件

得到满足，就应恢复交互，以保持参与度。

我们在有限状态机和行为树中实现了运动停止回退，将其作为对正在进行的交互的更高优先级中断。在 Shutter 运行期间，用户伸手靠近正在运动的机器人手臂可能构成潜在危险。为降低用户受伤的风险，机器人会暂停计划中的运动，直到用户离开。

与交互中途放弃类似，要在有限状态机中实现优先级回退，必须为每个其他状态定义一条转换至更高优先级状态的状态转换。随着表示交互所需的状态数量增加，所需的状态转换数量也会随之增加。此外，还必须定义从更高优先级状态返回原始状态的状态转换，以便交互能够恢复。在遍历和检查有限状态机时，保留原始上下文会带来额外开销。

相比之下，我们的行为树会持续检查优先级条件，以类似于检查用户是否在场的方式中断执行。如果优先级条件处于激活状态，运动停止就会继续阻止树的遍历。条件解除后，交互便会从同一位置继续进行，也就是从中断发生处继续。由于行为树的遍历可以通过控制流节点以及每个时钟周期事件对应的行为状态直接控制，因此恢复交互时无需像我们的有限状态机实现那样，另行记录中断时的上下文。

E. 行为树面临的挑战

交互中的某些组成部分，用有限状态机比用行为树更容易清晰地表达。由于行为树节点的设计是自包含的，即使是简单的动作模式也必须明确列出。因此，与设计简单的有限状态机相比，设计高效且正确的行为树需要更熟悉其架构。按钮按压就是行为树带来知识门槛的一个例子。在我们的有限状态机中，按钮按压直观地表示会改变机器人行为的状态转换。在我们的行为树中，每次按钮按压都被实现为并行节点的子节点。更复杂的用户输入形式，例如多种按钮按压组合或超时条件，则需要更复杂的结构。

V. 未来工作

行为树架构所确立的设计原则，为现实世界部署中的进一步工作提供了启发。具体而言，Shutter 的行为以及基于这些行为构建的树都是模块化且自成一体的，便于重复使用。得益于这种可复用性，我们能够快速开发并部署各种交互脚本。增强 Shutter 的感知能力，例如增加用于唤醒词检测的行为，或可实现更多交互模式。改进我们的群体检测方法，并将其融入更多交互中，则有望进一步提升交互的流畅度。例如，Shutter 可以询问附近哪些人应当被拍进照片，也可以请每位交互者分别查看照片。

通过定义并处理可靠的安全条件和故障回退机制，还可以改进多台 Shutter 的部署。安全地应对已知危险将减少对专家操作员或监督员的需求，从而提高自主性。

作为一个采用开源组件、成本相对较低的平台，Shutter 可以复制部署，以扩大真实场景中的数据收集和用户研究规模。由于行为树的架构有利于模块化，因此可以在公共空间中部署更多种类的不同交互。让 Shutter 担任群体调解者、社区信息提供者或面向公众的机器人系统教师，都是可以通过行为树描述的交互示例，并且能够复用我们许多现有的软件组件。

参考文献

[1] D. Bohus、C. W. Saw 和 E. Horvitz，“Directions Robot：真实场景中的体验与经验教训”，载于*2014* 年自主智能体与多智能体系统国际会议论文集，AAMAS ’14 系列。南卡罗来纳州里奇兰：自主智能体与多智能体系统国际基金会，2014 年，第 637–644 页。[2] M. Swofford、J. Peruzzi、N. Tsoi、S. Thompson、R. Mart´ın-Mart´ın、S. Savarese 和 M. V´azquez，“通过 Dante 提升社交意识：用于聚类对话交互者的深度亲和网络”，*ACM* 人机交互论文集，第 4 卷，CSCW1 期，第 1–23 页，2020 年。[3] J. G. Trafton、L. M. Hiatt、A. M. Harrison、F. P. Tamborello、S. S. Khemlani 和 A. C. Schultz，“ACT-R/E：用于人机交互的具身认知架构”，人机交互期刊，第 2 卷，第 1 期，第 30–55 页，2013 年 2 月。 [在线]。可访问：https://doi.org/10.5898/JHRI.2.1.Trafton [4] L. M. Hiatt、C. Narber、E. Bekele、S. S. Khemlani 和 J. G. Trafton，“用于人机协作的人类建模”，机器人学研究国际期刊，第 36 卷，第 5–7 期，第 580–596 页，2017 年。 [ ] 10 1177 0278364917690592 在线。可访问：https://doi.org/ 。 / F. Del Duchetto、P. Baxter 和 M. Hanheide，“Lindsey 导览机器人——博物馆长期部署中的使用模式”，载于*2019* 年第 *28* 届 *IEEE* 机器人与人类交互通信国际会议（*RO-MAN*）。IEEE，2019 年，第 1–8 页。[6] V. de Araujo、A. P. G. Almeida、C. T. Miranda 和 F. de Barros Vidal，“一种用于搜索与救援任务无人机控制的并行分层有限状态机方法”，载于*2014* 年第 *11* 届信息学、控制、自动化与机器人国际会议（*ICINCO*），第 1 卷。IEEE，2014 年，第 410–415 页。[7] A. Bauer、D. Wollherr 和 M. Buss，“用于人机通信的信息检索系统——询问方向”，载于*2009* 年 *IEEE* 国际机器人与自动化会议，2009 年，第 4150–4155 页。[8] T. Kanda、M. Shiomi、Z. Miyashita、H. Ishiguro 和 N. Hagita，“购物中心中的通信机器人”，*IEEE* 机器人学汇刊，第 26 卷，第 5 期，第 897–913 页，2010 年。[9] S. Wang 和 H. I. Christensen，“TritonBot：长期自主运行导览机器人部署的初步经验教训”，载于*2018* 年第 *27* 届 *IEEE* 机器人与人类交互通信国际研讨会（*RO-MAN*），2018 年，第 158–165 页。[10] M. Colledanchise 和 P. ¨Ogren，“机器人学与人工智能中的行为树：简介”，*ArXiv*，第 abs/1709.00084 卷，2017 年。[11] M. Iovino、E. Scukins、J. Styrud、P. ¨Ogren 和 C. Smith，“机器人学与人工智能中的行为树综述”，2020 年。[12] E. Coronado、X. Indurkhya 和 G. Venture，“机器人遇见儿童：面向真实场景中儿童—机器人交互的半自主控制系统开发”，载于*2019* 年 *IEEE* 第 *4* 届高级机器人与机电一体化国际会议（*ICARM*），2019 年，第 360–365 页。