



# 基于 Bots 的人机交互界面范式

王慧<sup>1,3</sup>, 姚乃明<sup>2,4</sup>, 董健<sup>5</sup>, 朱频频<sup>1,3</sup>, 陈辉<sup>2,4\*</sup>

1. 上海智臻智能网络科技股份有限公司小 i 机器人研究院, 上海 201803
  2. 中国科学院软件研究所人机交互北京市重点实验室, 北京 100190
  3. 上海智臻智能网络科技股份有限公司上海智能在线服务机器人工程技术研究中心, 上海 201803
  4. 中国科学院大学, 北京 100049
  5. 公安部第三研究所, 上海 200031
- \* 通信作者. E-mail: chenhui@iscas.ac.cn

收稿日期: 2017-10-30; 接受日期: 2018-01-17; 网络出版日期: 2018-04-08

国家重点研发计划项目 (批准号: 2016YFB1001201, 2017YFC0113100)、国家自然科学基金项目 (批准号: 61232013, 61422162, 61661146002)、中国科学院前沿科学重点研究计划 (批准号: QYZDY-SSW-JSC041)、上海市经济和信息化委员会专项资金 (批准号: 201602024) 和网络安全公安部重点实验室开放课题 (批准号: C16609) 资助项目

**摘要** 以聊天机器人为代表的 Bots 带动了以用户为中心的全新交互体验. 基于多通道自然语言交互的 Bots 不再依赖于桌面的图形环境, 源于桌面隐喻的传统 WIMP (windows, icon, menu, pointer) 范式也不再适用于基于 Bots 的人机交互, 因此需要建立一种新的支持移动交互的界面范式来规范界面的描述. 为此, 本文对化身隐喻 (avatar metaphor) 进行扩展, 提出了基于化身交互隐喻的 ASLI (avatar, scenario, language, instrument) 界面范式, 并给出了基于 ASLI 范式的界面实例和交互框架, 为基于 Bots 的人机交互系统提供了界面设计的指导. 最后, 讨论了未来 ASLI 范式的平台化发展趋势和衍生的生态系统.

**关键词** Bots, 人机交互, 界面范式, 化身隐喻

## 1 引言

有关 Bots 的定义和描述有很多, 从广义上可以认为所有能够自动完成用户任务的工具、应用和程序等都可以称为 Bots. 然而这种广义的 Bots 在很长时间内只是机械原理的呈现, 并不具备人机交互的条件. 计算机的发明促使信息化与自动化结合, 一方面可以通过程序控制和管理自动化的过程, 另一方面也开启了人机交互的研究和发展. 现在更加普遍的观点认为, 基于人类自然语言交互的、能够为用户执行任务的程序或软件称为 Bots<sup>[1]</sup>, 其代表就是对话式 Bots (conversational Bots), 其中包含了聊天机器人 (chatbots)<sup>[2]</sup> 和虚拟助手 (virtual assistant) 等概念<sup>[3]</sup>. 基于自然语言交互的 Bots 不但能提高人机交互效率, 还能大大降低商业成本, 因而广泛地应用在金融、电子商务、电子政务等行

**引用格式:** 王慧, 姚乃明, 董健, 等. 基于 Bots 的人机交互界面范式. 中国科学: 信息科学, 2018, 48: 475-484, doi: 10.1360/N112017-00210  
Wang H, Yao N M, Dong J, et al. Interface paradigm of Bots-based human-computer interaction (in Chinese). Sci Sin Inform, 2018, 48: 475-484, doi: 10.1360/N112017-00210

业领域中, 有效地提供智能客服、数据分析、流程管理等服务并获得了巨大的成功<sup>[2]</sup>, 而 Bots 也成为新型人机交互的主流方式并引起广泛的重视<sup>[4]</sup>.

人机交互是一门研究人类所使用的交互式计算系统的设计、实施、评估, 以及相关主要现象的学科, 其目的是为了让计算系统更容易地为人类所使用<sup>[5,6]</sup>, 而用户界面则是实现人机交互的软件或硬件介质. 在过去的数十年中, 主要流行过 3 种用户界面<sup>[7]</sup>: 批处理方式 (batch interface)、命令行方式 (command-line user interface, CLIs) 和图形用户界面 (graphical user interface, GUI) 方式. 基于桌面隐喻提出的 WIMP (windows, icon, menu, pointer) 界面范式将图形用户界面带进了大众市场, 并推动了 PC 的繁荣发展. 如今, 尽管计算机的运算和处理速度大大提高, 但是用户与计算机之间的交互能力却没有得到相应的提高, 其中一个重要的原因就是缺少一个高效的人机交互界面<sup>[8]</sup>. 特别是移动互联网的普及带来移动计算的需要, 同时伴随着自然人机交互的出现诞生了基于 Bots 的全新交互框架. 这种全新的自然人机交互方式已经不再局限于图形用户界面的 WIMP 界面范式, 而需要一种全新的界面范式进行表述.

本文基于 Bots 的人机交互的特点扩展了化身交互的界面隐喻 (avatar metaphor)<sup>[9]</sup>, 并提出了基于 Bots 交互的 ASLI (avatar, scenario, language, instrument) 界面范式, 以及基于 ASLI 交互范式的应用实例和交互框架, 证明了 ASLI 界面范式能够在满足移动计算需求下全面地描述基于 Bots 的自然人机交互过程.

## 2 国内外现状

自 1979 年首个基于 WIMP 范式的 Alto 计算机诞生后<sup>[10]</sup>, 基于桌面隐喻的图形用户界面计算机逐渐盛行, 随后, Apple 公司的 Macintosh 操作系统<sup>[11]</sup>、Microsoft 的 Windows 和 Unix 操作系统的 Motif 窗口<sup>[12]</sup> 都使用了这一界面特征. 然而在 WIMP 界面流行的二十多年中, 其缺点也渐渐展现出来, 这主要体现在 WIMP 界面中输入方式的单一<sup>[7]</sup>, 任何超出点击 (pointing) 操作的交互语言都不能被 WIMP 范式所描述; 仅通过窗口的呈现方式很难自然地呈现复杂关系<sup>[13]</sup>, 复杂多维关系的标示需要多窗口环境的展示, 增加了用户的认知负担; 缺乏对交互环境的描述, 任何交互都没有利用交互环境的信息和知识, 默认的静态交互环境很难实现自然交互<sup>[14]</sup>.

20 世纪 90 年代起, 对新一代用户界面的研究也再度兴起. 1990 年, Green 和 Jacob<sup>[15]</sup> 提出了 Non-WIMP 用户界面, 用来描述没有使用桌面隐喻的用户界面. 1997 年, Dam<sup>[16]</sup> 提出了 Post-WIMP 用户界面, 该界面的特点是“至少包含了一项不基于传统 2D 交互组件的交互技术”. 1993 年, Nielsen<sup>[17]</sup> 对 Post-WIMP 深入地研究并提出了 Non-Command User Interfaces, 该研究认为以往的用户界面都是基于命令的界面, 而下一代用户界面是非命令式的, 即计算机可以根据用户的交互动作分析其意图从而执行相应的任务. 同时基于 Post-WIMP 环境下提出的笔式用户界面范式 PIBG, 模拟了人们日常的纸笔工作环境, 帮助人们在保持自然工作方式的同时可以充分利用强大的计算资源<sup>[7,8]</sup>.

新型用户界面的研究促进了大量新型的人机交互模式的涌现, 例如移动交互<sup>[18,19]</sup>、语音和多通道交互<sup>[20]</sup>、情感交互<sup>[21]</sup>, 以及虚拟和增强现实<sup>[22,23]</sup> 等. 这些新型交互模式不但允许多通道的交互数据和语言, 还借鉴了大量用户在自然物理环境的知识使得人机交互更为自然, 被称为自然用户界面 (natural user interface, NUI)<sup>[24]</sup>. NUI 融入了用户在日常生活中对非数字世界的既有知识, 从而使得用户在交互知识、交互环境和交互方式上都能拥有更自然和直观的体验<sup>[25]</sup>. 因此, 相对于传统用户界面, NUI 具有颠覆性的改变. 2006 年, Jacob 等<sup>[26~28]</sup> 提出了基于真实的交互 (reality-based interaction, RBI) 概念, 归纳了 NUI 的共同特征. 在 RBI 框架体系下, NUI 包含了物理感知、身体意识和技能、环

境意识和技能, 以及社会意识和技能的 4 个层级的真实感. 该框架不但为 NUI 新型界面的特征和评估提供指导<sup>[25]</sup>, 更对新型自然人机交互的描述具有重要意义.

Bots 的发展与人机交互技术的发展并行, Massachusetts Institute of Technology 的 Weizenbaum 教授<sup>[29]</sup> 在 1966 年创造出世界上第一个基于自然语言处理的聊天机器人, 并取名为 ELIZA, 从此开启了聊天机器人的序幕. 以 ELIZA 为代表的聊天机器人主要通过文本或语音等方式来实现人机对话式交互, 用于模拟真人在对话中的行为. ELIZA 的诞生引起了人们的浓厚兴趣, 这是因为相比于冷冰冰的图形界面交互, 人们还是更倾向于基于自然语言的交互形式. 可以说 ELIZA 所展现的对话交互形式为后来各式各样的人机对话系统带来了灵感. 然而, 受限于数据的传输, 此时的 Bots 只能基于单机的交互模式完成简单的数据的检索、处理和分析. 直到 20 世纪 90 年代, 互联网的出现使得计算机可以通过一定的标准协议形成能够覆盖全球的信息传输和共享系统, 这直接促进了基于网页和基于 PC 端即时消息 (instant messaging, IM) 软件<sup>[30]</sup> 的聊天机器人的诞生, 并发展为较为成熟的商业应用, 如: 在线问答技术、虚拟聊天机器人、虚拟客服和消息平台等. 特别是在 21 世纪初, 移动互联网兴起、宽带无线接入技术的发展以及移动智能终端设备的普及, 使得人们可以随时随地地享受互联网带来的便捷服务和信息. 即时通信工具逐渐由 PC 端转入移动端, 且向着以平台为渠道、以用户为中心的个性化方向发展. 2011 年 Apple 发布虚拟助手 Siri<sup>[31]</sup>, 用户可以通过自然语音与手机的应用进行交互, 此后相似的智能助手纷纷涌现, 包括 Microsoft Cortana, Google Now<sup>[32]</sup>, Google Assistant, Amazon Alexa<sup>[33]</sup>, 三星 S Voice 等, 这些基于自然语言交互的智能助手开启了 Bots 应用的新时代. Microsoft 前 CEO Bill Gates 曾说: “机器人、无处不在的屏幕、语音交互, 这些都将改变我们看待‘电脑’的方式. 一旦看、听、阅读能力得到提升, 你就可以以新的方式进行交互.” 现在用户可以通过手机、平板电脑、智能家居和智能穿戴设备等多种渠道, 使用文本、语音、图片、视频等多通道的自然交互方式与计算机或其他用户进行交流, 基于 Bots 的自然交互正逐渐成为人机交互的主要方式.

### 3 基于 Bots 的界面范式

界面隐喻和界面范式是用户界面的重要组成部分. 一般来说, 人机交互系统所蕴含的人机交互界面的可被广泛接受的抽象含义称之为界面隐喻<sup>[34]</sup>, 如图形用户界面中的桌面隐喻. 界面范式是基于界面隐喻的技术描述, 是进行界面和交互设计的主导思想和指导方式<sup>[8]</sup>, 可以规范用户界面的设计和使用.

Bots 的兴起带动了以用户为中心的全新交互形式. 这主要体现在, 基于 Bots 的人机交互不再依赖于桌面的图形环境, 而是以自然语言为代表的多通道信息交互. 这意味着基于 Bots 的交互不再局限于图形界面的 WIMP 交互范式, 因而需要通过一种全新的界面交互范式进行描述. 在此, 首先给出基于 Bots 交互的界面隐喻, 并提出一种全新的基于 Bots 交互的 ASLI 界面范式, 即 “avatar, scenario, language, instrument”, 最后给出基于 Bots 的人机交互的 ASLI 界面范式示例.

#### 3.1 化身交互隐喻

在基于 Bots 的人机交互中, 用户与 Bots 之间一般通过自然语言来传递和表达信息. 传统基于图形界面的桌面隐喻 (desktop metaphor) 和笔式用户界面的笔纸隐喻 (pen/paper metaphor)<sup>[7,8]</sup> 显然不再适用. 因此本文提出一种基于化身的隐喻 (avatar metaphor). 这里所说的化身 (avatar) 是指用户或计算机的一种用于交互的社会表征<sup>[9]</sup>, 它建立了人与计算机之间信息交互的桥梁, 既可以是如同实体

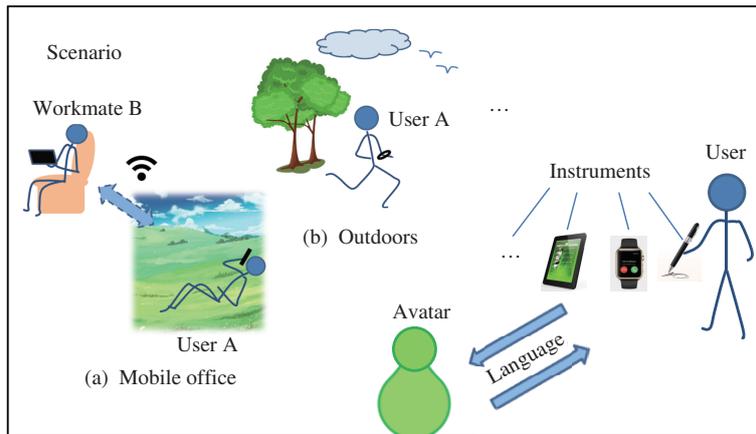


图 1 (网络版彩图) 基于 Bots 的 ASLI 交互范式示意图  
 Figure 1 (Color online) schematic diagram of ASLI paradigm

机器人或虚拟人像的可见形态,也可以是如虚拟助手的不可见形态.在化身隐喻下,用户可以通过包括文本、语音、符号、动作、表情等多通道数据进行人机交互,我们将所有这些可以用于交互信息的载体作为广义语言.相比于桌面隐喻和笔纸隐喻,基于化身的隐喻更适合以用户为中心的自然人机交互,可以将广义语言所覆盖的用户感知高度统一,从而实现更自然更高效的交互界面.

### 3.2 ASLI 界面范式

在上述基于化身的隐喻基础上,本文提出了一种全新的 ASLI 界面范式,该范式主要由 4 个成分构成: avatar, scenario, language 和 instrument,如图 1 所示.

avatar 是承载信息交互的载体,同时作为从用户获取输入信息以及向用户呈现输出信息的载体,是用户在交互过程中始终关注的交互对象.根据应用需求,avatar 既可以是可见形态,也可以是非可见形态.与多智能体 (multi-agent) 交互系统相似,每个 avatar 有自身的表示、抽象和控制的 PAC (presentation, abstract, control) 机制<sup>[35]</sup>.基于 avatar 的交互能够通过各种多通道数据来平衡用户与计算机之间的信息交流,每个 avatar 都有自身的表示 (presentation),可以更有针对性地凸显其操作,具有良好的功能可见性 (affordance),这些为以用户为中心的自然人用户界面奠定基础,同时提升系统的开发效率.

scenario 描述的是包含 avatar 和用户在内的交互环境,可以随着时间和空间的变化而动态变化,主要描述应用场景、环境变化,以及上下文对交互语义理解的影响.scenario 所包含的内容体现了交互所涉及的局部世界中所有可以被 instrument 捕捉到的信息,以及该局部世界所特有的知识描述,如时间、地点、自然景物、人物、对白等.scenario 是 Bots 范式与以往 WIMP 范式和 PIBG 范式都不同的地方,是以用户为中心的自然人机交互界面所特有的,同时也是移动计算的必然要求.在以用户为中心的设计导向下,用户的心理、广义语言以及其中隐含的语义信息都会随着外部环境而改变,而随着移动计算的需求这种外部环境的影响逐渐增强.为了减少自然交互中用户与计算机之间的执行和评估的鸿沟,环境和场景的因素必然成为交互界面需要考虑的一部分.在移动计算中,scenario 的转换一般由 avatar 通过将用户的输入作为触发或判断依据完成的,在这个过程中用户可以主动地进行 scenario 切换,也可以非主动地由 avatar 进行判断.

language 描述交互信息的载体和交互方式,包括语音、文本、表情、动作等可以传递交互信息的多通道广义语言.也就是说用户将通过广义语言与 avatar 传递信息和发送命令,而 avatar 也将通过广义语言向用户呈现输出反馈.但是并非所有广义语言都能用来与 avatar 进行交互,这是因为与 avatar 交互的目的是为了通过功能接口调用应用逻辑完成相关任务,与应用的功能相对应的是一系列交互原语集合,每个交互原语都对应着明确的应用功能或执行操作.因此,这里的 language 是受限的,只有那些含有交互信息的,能与交互原语匹配的广义语言才能用来与 Bots 进行交互.用户通过 language 进行人机交互的时候,其关注点始终保持在执行目标上,即以任务为导向,而非考虑执行任务的过程和操作.在图形界面的 WIMP 范式中,用户注意力可能会向 icon/menu 的交互组件发生转移,这可能给用户的认知过程带来难度从而降低交互效率.ASLI 范式摒弃了 WIMP 范式中的 icon/menu 和 PIBG 范式中的 icon/button 组件,这是因为这些组件一般隐藏在 language 的语义操作中,而非在图形界面中有明确的显式表达.于是,WIMP 范式中通过 icon/menu 进行 pointing 操作的过程,在 ASLI 范式中只需通过 language 进行任务导向的交互来完成,这其实体现了用户在纵向操作层次的精简,降低了用户的认知要求,从而提高人机交互的操作效率.设计良好的 language 输入可以让用户与计算机的交互接近于人人交互一般自然和流畅.

instrument 描述用户与 avatar 交互所使用的交互工具和设备,包括探测各类广义语言的传感器和辅助工具设备,如探测声音的麦克风、捕捉动作的摄像机、感知手势的触摸屏和笔等.在与 Bots 的交互中,instrument 可以直接将用户的操作转化为明确的指令和意图,这种转化不会随着 scenario 的变化而变化,因此用户通过 instrument 进行的交互行为是可以独立于 scenario 而直接标准化的.

### 3.3 基于 ASLI 范式的交互实例

为了更好地理解基于 ALIS 范式的人机交互,通过图 1 所示的“移动办公”和“户外”应用场景示例来进一步说明.用户首先在移动办公的 scenario 中,如图 1(a) 所示,该 scenario 包含用户 A、同事 B、avatar、书桌、电脑等,其中 avatar 是用户 A 的虚拟助理.用户 A 可以通过智能手机或平板电脑等智能终端的 instruments 与 avatar 进行交互.用户 A 可以通过语音 language 唤醒 avatar,询问:“今天下午我有什么日程安排?” avatar 通过麦克风 instrument 捕捉到用户的语音数据作为 language,然后进行语义分析和意图理解,并通过功能接口调用日程应用,该应用将根据用户个人信息知识将当日日程数据通过功能接口返回 avatar. avatar 通过语音 language 将该日程播报给用户 A:“今天下午两点钟有与客户的会议,地点在一楼会议室.今天晚上七点钟跑步 5 公里.”用户 A 使用语音 language 命令:“将客户会议的安排发送邮件给同事 B.”该语音 language 通过麦克风 instrument 传递给 avatar 进行语义分析和意图理解,然后通过功能接口调用日程和邮件应用,应用将根据用户个人信息知识和办公室 scenario 的知识,找到同事 B 的邮箱地址,将日程发送至该地址,并将发送成功的提示回馈给 avatar, avatar 通过语音 language 呈现给用户:“将下午两点与客户的会议日程发送邮件给同事 B,发送成功.”在这个场景示例下,用户主要通过语音 language 与虚拟助手进行交互,此时虚拟助手等价于语音助手,ASLI 范式可以直接应用在语音助手的人机交互中.

傍晚,用户 A 在户外跑步,如图 1(b) 所示,此时 scenario 转换到户外运动,所包含的内容变为:用户 A, avatar, 周围的自然景物、人物、动物等,用户 A 可以通过智能手机或智能穿戴设备与 avatar 进行交互,智能穿戴设备上的 avatar 可以是不可见的形态,也可以通过手机屏幕的虚拟人像等方法呈现可见形态.用户首先用语音 language 通过麦克风 instrument 唤醒 avatar,并命令:“进入运动模式,记录运动里程和心跳参数.” avatar 分析语义得到用户的意图为“开启运动模式”、“记录运动数据”、“记录生理参数”,并通过功能接口分别调用模式应用、GPS 应用和生理信号检测应用等.所调用的应用

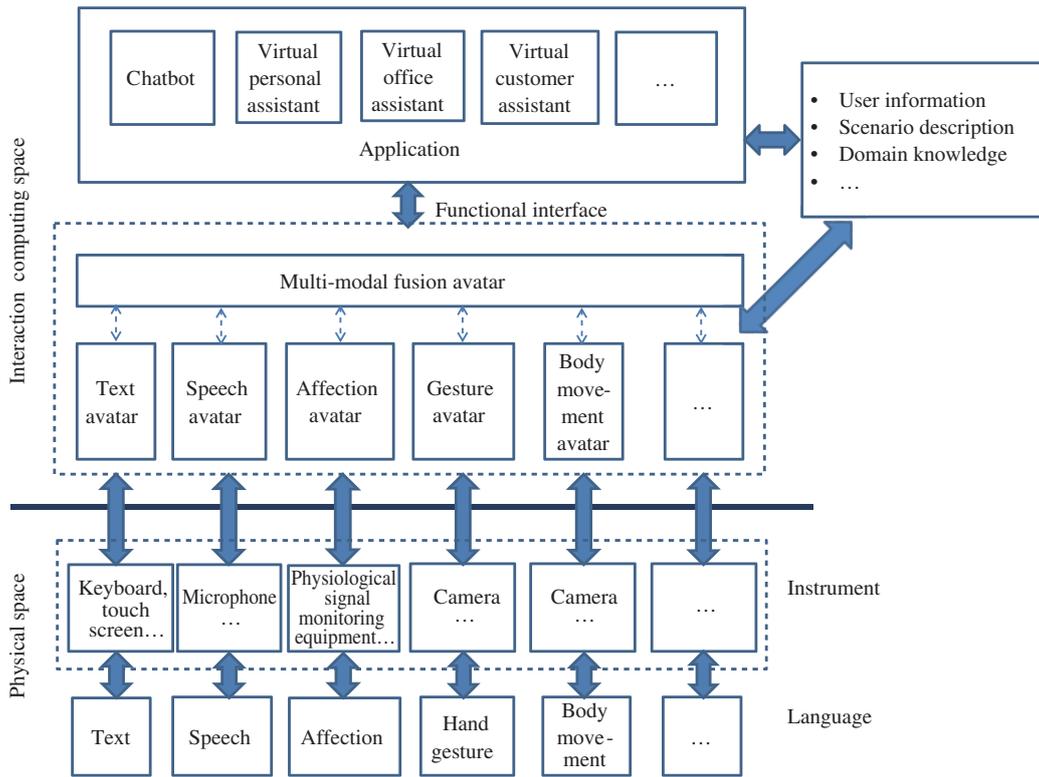


图 2 (网络版彩图) 基于 Bots 的交互框架  
 Figure 2 (Color online) Bots-based interaction framework

根据相应的逻辑开启运动模式并执行相关设置; 通过 instrument, 即 GPS 定位设备, 来追踪用户跑步的路线计算运动里程和速度; 通过心跳检测设备 etc instrument 来捕捉用户的生理数据 language, 并根据用户以往心跳数据、运动记录等个人信息, 以及户外运动 scenario 的知识, 通过 avatar 给出实时反馈. 例如通过语音 language 播报: “您今日跑步里程已达 3 公里, 完成目标的 60%.” 或 “您当前的心跳超过 150, 请注意调整呼吸和运动强度.” 完成运动的时候, 用户 A 通过手机摄像头 instrument 拍照户外景色, 将图片 language 发送给 avatar, avatar 将调用图像理解应用, 应用根据用户个人信息, 户外运动 scenario 的知识, 将图像理解的内容传递给 avatar, avatar 根据用户当前的 scenario, 可以通过文本 language 回复: “景色真美!”, 以及虚拟人像的表情 language 发送笑脸表情来呈现喜悦的感情. 在户外运动的场景示例中, 用户可以通过语音、生理信号、图像和文本等多种 language 与虚拟助手进行交互. 因此, ASLI 范式不再局限于语音助手, 还可以应用在基于 Bots 的更广泛的人机交互中.

#### 4 基于 ASLI 范式的交互框架

基于 ASLI 范式的交互框架如图 2 所示, 其具体交互过程为如下.

首先, 在物理空间的用户可以使用文本、语音、手势、动作等通道或者多通道融合 language, 可以通过不同类型的 instrument 将交互信息传递给交互计算空间的 avatar. 随着 scenario 的变化, 用户的 language 可能发生通道、语义等内容的变化, 而 instrument 的操作和执行则不会变化.

其次, 根据交互数据和应用需求, 相应的 avatar 或多通道融合 avatar 将根据用户信息、场景描述

和领域知识等 scenario 的信息, 将用户的输入 language 进行语义分析, 并通过功能接口来调用相应的应用. 这里的 language 必须与交互原语匹配才能调用后端应用, 不能匹配交互原语的内容无法进行交互.

最后, 被调用的应用将根据用户信息、场景描述和领域知识等 scenario 的信息, 基于应用逻辑再度通过功能接口将交互信息传递给相应的 avatar, 由 avatar 向用户呈现不同通道的信息反馈.

由上述的交互框架可以看出, 基于 Bots 交互的语义分析和应用逻辑都受到包括用户信息、场景描述和领域知识等有关 scenario 信息的影响, 将这些影响因素统称为 scenario 知识. 这种知识描述的内容十分广泛, 包含了对 Bots 在应用场景中所需解决的全部问题的表述, 比如用户信息可以用来对 Bots 的个性化知识进行建模, 而应用场景和领域信息可以用来对不同场景应用和领域知识进行建模, 除此之外还可能多种信息都可以用来构建知识从而对 Bots 的操作执行产生影响. 基于人类自然语言的特点, 这种以用户为中心的知识表述一般是以用户意图或任务为导向的, 并且可以为多个 Bots 所共享. 这种知识的共享构成了 Bots 之间的交互能力, 是 Bots 之间彼此协作的一种重要模式, 这主要体现在不同 Bots 可能基于同一个知识体系与用户进行交互.

从图 2 的交互框架还可以看到, 用户与一个 avatar 进行交互, 完全可能从另一个 avatar 得到反馈呈现, 这体现了 Bots 之间协作的另一种重要方式. 这主要体现在当用户的意图对单一 Bots 的操作过于复杂或难以回馈时, 可以通过多个 Bots 的协作来实现交互过程. Bots 的这种协作模式可以覆盖多通道的用户输入, 从而平衡人机交互中信息传递的对等性.

## 5 ASLI 范式的平台化与生态系统

21 世纪初, 随着互联网进入了 Web 2.0 的时代, 以平台为渠道、以用户为中心的个性化需求和互动成为了互联网的核心. 目前, 微信、Facebook, Apple, Google, Amazon 等公司都发布了 Bots 平台并对开发者提供各种 API<sup>[1]</sup>. 微软还提出了“对话即平台 (conversation as a platform, CaaP)”的理念, 改变了企业和用户之间的交流方式. 全球最具权威的 IT 研究与顾问咨询公司 Gartner 指出<sup>[36]</sup>: “会话式人工智能平台 (conversational AI platforms) 将是信息技术的下一次重大范式转变”.

由此可见, 基于 Bots 的新型人机交互模式已成为平台化服务的重要组成部分, 而 ASLI 作为 Bots 的人机交互界面范式也将逐步走向平台化. ASLI 平台化过程可能包含多种实现层面, 例如 ASLI 范式的交互框架将与 Bots 平台的开发框架相统一; 为了满足广泛的开发和应用需求, Bots 平台的对外接口将逐步形成统一标准规范, 这对于 ASLI 的 language 和 instrument 的接口规范有重要指导意义; 为了根据实际需要进行开发, Bots 平台可能划分不同的应用领域和功能服务, 这对于 scenario 模型的平台化构建有重要指导意义.

ASLI 范式的平台化十分重要且必要, 不但可以规范化新型人机交互和界面设计, 还将从交互对象、交互环境、交互语言和交互设备等多个层面为 Bots 的开发提供便捷的技术环境和功能接口. 在这样平台化的大环境下, 构建 Bots 的过程将变得十分便捷, 开发者无需担心系统集成、编程, 以及维护的问题, 能更加专注于业务功能本身的开发. 一些平台提供的自学习能力还可以实现 Bots 功能的自动扩展升级, 使其架构随着使用过程更加趋于稳定, 运维管理也更加便利.

稳定且规范化的 ASLI 平台架构也将进一步促进 Bots 之间的协作以及生态系统的发展. 在未来的应用场景中, Bots 可以在数据、计算资源、算法、技术架构、功能模块和领域应用等不同层面进行扩展和协作. 当一个 Bots 无法回答用户问题的时候, 它可以主动向另一个 Bots 寻求帮助来获取用户希望的解答, Bots 仿佛具有人类智慧一般进行协同工作. Bots 这种多层面的协作将衍生出基于 Bots 的

生态系统, 使得人机交互与人工智能在多层面得到有机的结合. 基于 Bots 的自然语言交互是人工智能在应用上的重要体现, 而人工智能的快速发展也为人机交互提供以大数据和计算能力为基础资源, 以机器学习为基础算法, 结合自然语言处理、知识工程、语音技术和计算机视觉等人工智能的核心技术, 并通过功能接口支持上层各领域中 Bots 的产业化应用以及物联网等多层面的生态系统.

## 6 总结

Bots 以其独特的自然人机交互方式已成为衔接用户与服务的主流交互接口. 基于 Bots 的新型人机交互特点, 本文提出了化身交互的界面隐喻以及一种全新的 ASLI 界面范式, 其中 avatar 是信息收集和呈现的载体, scenario 描述包括用户和 avatar 在内的外部环境的影响, language 描述交互的方式, instrument 描述用户与 avatar 进行交互的设备和工具. ASLI 交互范式充分考虑了以用户为中心的自然语言人机交互的特点, 并且满足移动式计算的需求, 为基于 Bots 的人机交互系统提供了交互界面设计的指导. 随着 Bots 平台化的发展, ASLI 范式的平台化将促进 Bots 之间的协作和相应生态系统的发展.

## 参考文献

---

- 1 Bruner J, Barlow M. What Are Conversational Bots? Sebastopol: O'Reilly Media Inc, 2016
- 2 Backer V L, Revang M. Four Use Cases for Chatbots in the Enterprise Now. Gartner Report, ID: G00316146, 2017
- 3 Brant K F, Austin T. Hype Cycle for Artificial Intelligence. Gartner Report, ID: G00314732, 2017
- 4 Andrews W. Where You Should Use Artificial Intelligence — and Why. Gartner Report, ID: G00328113, 2017
- 5 Dong J M, Fu L M, Salvendy G. Human-Computer Interaction: User Centered Design and Evaluation. Beijing: Tsinghua University Press, 2002 [董建明, 傅利民, Salvendy G. 人机交互: 以用户为中心的设计和评估. 北京: 清华大学出版社, 2002]
- 6 Dix A, Finlay J, Abowd G, et al. Human-Computer Interaction. Upper Saddle River: Prentice-Hall, 1993
- 7 Tian F, Mu S, Dai G Z, et al. Research on a pen-based interaction paradigm in Post-WIMP environment. Chinese J Comput, 2004, 27: 977-984 [田丰, 牟书, 戴国忠, 等. Post-WIMP 环境下笔式交互范式的研究. 计算机学报, 2004, 27: 977-984]
- 8 Dai G Z, Tian F. Pen-based User Interface. Hefei: Press of University of Science and Technology of China, 2008 [戴国忠, 田丰. 笔式用户界面. 合肥: 中国科学技术大学出版社, 2008]
- 9 Meadows M. I, Avatar: the Culture and Consequences of Having a Second Life. Thousand Oaks: New Riders Press, 2007
- 10 Thacker C P, McCreight E M, Lampson B W, et al. Alto: a personal computer. In: Computer Structures: Readings and Examples. New York: McGraw-hill, 1979. 549-572
- 11 Williams G. The apple macintosh computer. Byte, 1984. <http://www.1000bit.it/support/articoli/apple/mac128.pdf>
- 12 Zhu X F, Gu Y L, Cao Y. Study of graphical user interface development technique based on Motif in UNIX. J South Yangtze Univ (Nat Sci Edit), 2002, 1: 38-41 [朱雪峰, 顾耀林, 曹莹. UNIX 平台基于 Motif 环境的图形用户界面开发技术. 江南大学学报 (自然科学版), 2002, 1: 38-41]
- 13 Koike H. The role of another spatial dimension in software visualization. ACM Trans Inf Syst, 1993, 11: 266-286
- 14 Lv F. Research on reality-based interaction techniques and applications. Dissertation for Ph.D. Degree. Beijing: University of Chinese Academy of Sciences, 2012 [吕菲. Reality-based interaction 交互技术及应用研究. 博士学位论文. 北京: 中国科学院大学, 2012]
- 15 Green M, Jacob R. Siggraph'90 workshop report: software architectures and metaphors for non-wimp user interfaces. SIGGRAPH Comput Graph, 1991, 25: 229-235
- 16 Dam A V. Post-wimp user interfaces. Commun ACM, 1997, 40: 63-67
- 17 Nielsen J. Noncommand user interfaces. Commun ACM, 1993, 36: 82-99
- 18 Satyanarayanan M. Pervasive computing: vision and challenges. IEEE Pers Commun, 2001, 8: 10-17

- 19 Saha D, Mukherjee A. Pervasive computing: a paradigm for the 21st century. *Computer*, 2003, 36: 25–31
- 20 Waibel A, Vo M T, Duchnowski P, et al. Multimodal interfaces. *Artif Intell Rev*, 1996, 10: 299–319
- 21 Picard R W. *Affective Computing*. Cambridge: the MIT Press, 1997
- 22 Azuma R, Bailiot Y, Behringer R, et al. Recent advances in augmented reality. *IEEE Comput Graph Appl*, 2001, 21: 34–47
- 23 Azuma R T. A survey of augmented reality. *Presence-Teleop Virt Environ*, 1997, 6: 355–385
- 24 Wigdor D, Wixon D. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. San Francisco: Morgan Kaufmann, 2011
- 25 Lv F, Tian F, Du Y, et al. NEM: a natural user interface evaluation method based on reality framework. *J Comput-Aided Des Comput Graph*, 2017, 29: 2076–2082 [吕菲, 田丰, 杜一, 等. NEM: 基于真实感框架的自然用户界面评估方法. *计算机辅助设计与图形学学报*, 2017, 29: 2076–2082]
- 26 Jacob R J K. What is the next generation of human-computer interaction? In: *Proceedings of CHI'06 Extended Abstracts on Human Factors in Computing Systems*, Montréal, 2006. 1707–1710
- 27 Jacob R J K, Girouard A, Hirshfield L M, et al. Reality-based interaction: unifying the new generation of interaction styles. In: *Proceedings of CHI'07 Extended Abstracts on Human Factors in Computing Systems*, San Jose, 2007. 2465–2470
- 28 Jacob R J K, Girouard A, Hirshfield L M, et al. Reality-based interaction: a framework for post-wimp interfaces. In: *Proceeding of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*, Florence, 2008. 201–210
- 29 Weizenbaum J. ELIZA — a computer program for the study of natural language communication between man and machine. *Commun ACM*, 1966, 9: 36–45
- 30 Nardi B A, Whittaker S, Bradner E. Interaction and outeraction: instant messaging in action. In: *Proceedings of ACM Conference on Computer Supported Cooperative Work*, Philadelphia, 2000. 79–88
- 31 Allworth J. Apple's Siri is as revolutionary as the Mac. 2011. <https://hbr.org/2011/10/apples-siri-is-as-revolutionar>
- 32 Thakur S. Personalization for Google now: user understanding and application to information recommendation and exploration. In: *Proceedings of the 10th ACM Conference on Recommender Systems*, Boston, 2016
- 33 López G, Quesada L, Guerrero L A. Alexa vs. Siri vs. Cortana vs. Google assistant: a comparison of speech-based natural user interfaces. In: *Proceedings of International Conference on Applied Human Factors and Ergonomics*, Los Angeles, 2017. 241–250
- 34 Koved L, Selker T. Room with a view (RWAV): a metaphor for interactive computing. 2007. <http://www.research.ibm.com/people/k/koved/papers/rwavrc.pdf>
- 35 J Coutaz. PAC, an object oriented model for dialog design. In: *Proceedings of the 2nd IFIP Conference on Human-Computer Interaction*, Stuttgart, 1987. 431–436
- 36 Burke B, Austin T, Searle S, et al. Top 10 Strategic Technology Trends for 2017: Conversational Systems. Gartner Report, ID: G00319579, 2017

## Interface paradigm of Bots-based human-computer interaction

Hui WANG<sup>1,3</sup>, Naiming YAO<sup>2,4</sup>, Jian DONG<sup>5</sup>, Pinpin ZHU<sup>1,3</sup> & Hui CHEN<sup>2,4\*</sup>

1. *Xiaoi Research, Shanghai Xiaoi Robot Technology Co., Ltd, Shanghai 201803, China;*

2. *Beijing Key Lab of Human-Computer Interaction, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China;*

3. *Shanghai Engineering Research Center of Intelligent Service Robot, Shanghai Xiaoi Robot Technology Co., Ltd, Shanghai 201803, China;*

4. *University of Chinese Academy of Sciences, Beijing 100049, China;*

5. *The Third Institute of Ministry of Public Security, Shanghai 200031, China*

\* Corresponding author. E-mail: [chenhui@iscas.ac.cn](mailto:chenhui@iscas.ac.cn)

**Abstract** Bots, usually known as chatbots, have given rise to a brand-new pattern of user-centered interaction. Owing to the development of multi-modal natural language interaction, Bots no longer rely on the graphical desktop for interaction. Consequently, a new interface paradigm is required to normalize the bots-based human-computer interaction, which is no longer limited to the WIMP (windows, icon, menu, and pointer) paradigm of the desktop paradigm. In this paper, a new ASLI (avatar, scenario, language, and instrument) interface paradigm is proposed on the basis of the extended avatar metaphor. The framework of the Bots-based interaction is provided as well as the corresponding interaction instances, which provide guidance for the interface design of Bots-based interaction. The platforms of Bots and ASLI are also discussed and are believed to be able to derive a Bots-based ecosystem in the near future.

**Keywords** Bots, human-computer interaction, interface paradigm, avatar metaphor



**Hui WANG** was born in 1984. She received her Ph.D. degree in optics from Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Science, and her Ph.D. degree in Physics from Université de Lille 1, France in 2013. Currently, she is a senior researcher at Shanghai Xiaoi Robot Technology Co., Ltd. Her research interests include computer vision, medical image reconstruction, affective computing, and human-computer interaction.



**Naiming YAO** was born in 1986. He received his M.S. degree in computer software and theory from Capital Normal University. Currently, he is a Ph.D. candidate at the Institute of Software, Chinese Academy of Sciences, Beijing and the University of Chinese Academy of Sciences, Beijing. His research interests include human-computer interaction, affective computing, machines learning, and computer vision.



**Pinpin ZHU** was born in 1976. He received his Ph.D. degree in physics from the Chinese Academy of Sciences, Shanghai in 2003. Currently, he is the founder and CEO of Shanghai Xiaoi Robot Technology Co., Ltd. His research interests include natural language processing, semantic analysis and understanding, knowledge engineering, and smart big data technology.



**Hui CHEN** was born in 1974. She received her Ph.D. degree in computer science from the Chinese University of Hong Kong, Hong Kong in 2006. Currently, she is a professor at the Institute of Software, Chinese Academy of Sciences. Her research interests include human-computer interaction, affective interaction, haptics, and virtual reality.