



智能时代的人机交互范式

张小龙^{1,2*}, 吕菲³, 程时伟⁴

1. 太原理工大学计算机科学与技术学院, 太原 030027, 中国

2. College of Information Sciences and Technology, Pennsylvania State University, University Park 16801, USA

3. 北京邮电大学数字媒体与设计艺术学院, 北京 100876, 中国

4. 浙江工业大学计算机科学与技术学院, 杭州 310023, 中国

* 通信作者. E-mail: zxl_psu@foxmail.com

收稿日期: 2017-11-02; 接受日期: 2018-01-29; 网络出版日期: 2018-04-08

国家自然科学基金 (批准号: 61572344, 61772468, 61303162) 和北京市社会科学基金 (批准号: 16YTC033) 资助项目

摘要 本文首先介绍范式的含义, 然后讨论人机交互领域中范式这一概念的使用. 以人机交互领域最有影响力的 WIMP 范式为例, 讨论界面范式所包含的主要成分, 并在此基础上分析了目前主要的智能系统的界面的特点和不足. 最后本文提出了针对智能系统界面设计的 RMCP 界面范式, 强调智能系统界面设计需要考虑系统的角色 (role)、交互模态 (modal)、交互命令 (commands) 和信息展示方式 (presentation style) 这几个基本因素.

关键词 人机交互, 界面范式, 智能系统

1 引言

智能计算机系统的快速发展对交互设计提出新的挑战. 智能系统中的人机界面设计往往采用语音、姿态等灵活、自然的模式, 简化了用户与系统的交互. 但学习和适应各类智能系统不同的用户界面也无疑会增加用户的认知负担, 影响用户体验. 如何为智能系统设计好的用户界面呢? 从人机交互领域的发展历史看, 好的用户界面设计往往依赖于某种界面范式, 例如个人计算机系统界面设计中的 WIMP (windows, icons, menus, a pointing device) 界面范式和笔交互系统中的 PGIS (paper, gadget, icon, sketch) 界面范式^[1].

迄今为止, 针对智能系统的界面范式的研究还是一个空白. 本文探索智能系统的界面范式, 力图为今后智能系统交互界面设计方法提供一些思路. 本文首先介绍对范式这一概念的不同解读, 然后通过 WIMP 界面范式的解析来讨论一个界面范式应该考虑的基本要素. 在此基础上, 进一步分析目前智能系统中常见的交互界面的特点, 并提出包含有角色 (role)、交互模态 (modal)、交互命令 (commands), 以及信息展示方式 (presentation style) 4 个基本要素的 RMCP 界面范式.

引用格式: 张小龙, 吕菲, 程时伟. 智能时代的人机交互范式. 中国科学: 信息科学, 2018, 48: 406-418, doi: 10.1360/N112017-00217
Zhang X L, Lyu F, Cheng S W. Interaction paradigm in intelligent systems (in Chinese). Sci Sin Inform, 2018, 48: 406-418, doi: 10.1360/N112017-00217

目前智能系统的种类和概念有很多,所以有必要首先界定本文所讨论的智能系统.这里的侧重点在交互层面,即用户是如何为系统提供信息输入、如何对系统的输出做出响应.多数交互系统都包含前端和后端两部分:前端负责接收用户的输入信息,并向用户展示相关的系统和结果信息;后端则根据用户的输入信息,产生用户需要的结果.一个交互系统的智能性可以体现在系统的前端、后端,或者二者兼而有之.前端的智能性往往表现为可以接受灵活、复杂的用户信息输入方式,例如语音、手势等方式.后端的智能性往往体现在对信息的处理和整合方面,如对照片的识别和分类、对文本数据的自动翻译等.一个后端智能的系统并不一定要求前端具有智能性,例如百度识图等智能图片识别系统依然依靠基于鼠标、键盘等传统前端交互方式.

本文针对的是具有前端智能性的系统,即一个能识别用户基于语音、手势等自然模态的输入信息,并能以某种方式将结果展现给用户的系统.很多后端智能系统需要用户的介入,以保证结果的有效性和可靠性.如何与这类系统有效地交互也是人机交互领域的重要研究问题,但这类问题的侧重点在于用户对智能系统后端的有效理解(如深度学习算法)和交互^[2].本文关心的问题则是智能的前端系统与用户有效交互的问题.

2 范式的基本含义

本节首先回顾范式这一概念的起源,然后介绍范式在人机交互领域中的不同含义.

2.1 范式在科学领域的起源和使用

范式由英文 paradigm 翻译而来.在科学领域内,范式这一概念的使用起源于美国科学哲学家 Thomas Samuel Kuhn 的《科学变革的结构》一书^[3,4].在该书中,Kuhn 把范式定义为一个学科中被广为接受的定理、应用、实验等基本科学实践,并指出每个学科都有其基本范式,而一个学科的变革往往伴随着基本范式的变革(paradigm shift).例如在物理学领域,研究关注点从牛顿力学到量子力学的变革伴随着研究方法从以牛顿 3 个定律和相关实验为核心的传统范式到以 Planck 和 Bohr 等的理论和相关实验为核心的新范式的转变.之后,“范式变革”在科学界被经常提及.

但对范式这一概念的使用和理解也比较混乱.虽然 Kuhn 的本意是用范式来指代一个学科中的典型例证,如成功的实验、方法、模型等,但这个着眼于微观层面的概念也被他人扩展到了宏观层面,用来指代一个学科所关心的基本科学问题和所使用的基本研究方法.例如 Gray^[5]用范式来指代科学研究的总体方法,并把科学研究的方法根据不同的历史阶段划分为基于实验的科学、基于理论的科学、基于计算的科学和基于数据的科学 4 个基本范式.

2.2 人机交互领域对范式的理解

在人机交互领域,范式这一概念也涉及宏观和微观两个层面的含义.

2.2.1 宏观层面的理解

宏观上,范式被用来指代人机交互领域所关心的基本问题以及基本研究方法. Harrison 等^[6]指出,人机交互研究经历了 3 个基本研究范式.第 1 个范式是以人因(human factors)理论为基础,研究着眼于与系统使用相关的物理设备属性,以改善交互设备的可用性为目的.第 2 个范式则以人类信息处理模型(human information processing)^[7]为基础,研究着眼于人与计算机交互时的认知过程,以提

高系统用户界面与人的认知特性, 以及任务特性的匹配性为目的. 第 3 个范式则超越了一般性的认知特性、任务特性的范畴, 强调用户的行为与特定的系统使用环境有关, 研究对象也包括那些用户和计算机系统之外的、与环境背景相关的因素.

不同的范式意味着不同的研究方法和相关领域. 第 1 个范式下的主要方法是系统设计, 而在第 2 个范式下, 研究方法强调系统设计和用户实验的并重. 第 3 个范式下的科学问题需要结合心理学、社会学等学科, 通过实地研究等手段全面理解用户、任务, 以及系统使用环境等因素的特点.

2.2.2 微观层面的理解

在人机交互领域, 范式这一词也被广泛地用来指代针对某个具体问题所涉及的框架、应用和方法等. 在这个层面, 范式的含义有多种. 例如在 Carroll^[8] 有关人机交互这一学科的阐述中, 范式就有不同的含义: 有时它被用来指代用户与信息系统的交互方式 (如基于浏览的交互范式和基于搜索的交互范式), 有时又被用来指代一个界面的设计方法 (如命令行范式), 有时又等同于一个理论模型 (如社会计算范式) 等.

总的来说, 范式这一概念与一个领域的一般性方法论有关, 更侧重于抽象性描述所涉及的问题和方法的主要成分, 但往往不涉及方法执行层面的具体实现形式.

3 界面范式

在人机交互领域中, 范式更多地被用来指代界面范式. 界面范式一般是指用户界面的设计模型或模式^[9,10]. 一个界面范式通常只定义界面设计中应该考虑的主要内容 (如交互部件), 而不涉及内容的具体功能、表现形式和实现方式. 例如 WIMP 界面范式虽然明确了一个界面应该包含代表窗口、图标、菜单、点击设备 4 个主要部分, 但并没有限定它们的具体设计和实现方式. 因此不同操作系统下基于 WIMP 范式的图形用户界面虽然都包含了这些基本设计元素, 但又有其各自的标准和特点.

3.1 WIMP 界面范式

WIMP 范式作为计算机用户界面中应用最为广泛的界面范式, 定义了设计一个图形用户界面时应该考虑的部件. 本小节解析 WIMP 范式中各个部件的作用, 然后总结用户界面应该具有的基本要素.

在人机交互领域内部, 研究人员对 WIMP 4 个部分所代表的具体内容曾有过不同的解释^[11], 但目前普遍为业界所接受的解释为 WIMP 的 4 个基本部件是窗口 (windows)、图标 (icons)、菜单 (menus), 和点击设备 (a pointing device)^[12]. 这里的讨论遵循这一解释.

3.1.1 窗口

一个窗口是具有边界的一个空间区域, 界定了一个应用程序所有相关信息在显示屏幕上应该出现的位置, 所以窗口可以被理解为一个应用程序的信息载体, 用于展示用户的输入信息 (如所编辑的文字)、系统反馈信息 (如运行结果) 等.

在支持多任务的系统中, 窗口还定义了一个用户命令的影响范围, 以确保一个命令被施加到所期望的应用或任务中. 例如当用户需要粘贴一段文字到记事簿时, 需要确保在记事簿窗口施加该粘贴命令, 否则粘贴任务就会失败.

3.1.2 图标和菜单

图标和菜单是交互过程中的交互对象. 和命令行交互范式相比, WIMP 中的图标和菜单通过显性的方式 (如图形化的图标或文字化的菜单选项) 来展现系统所接收的命令, 使得用户无需从记忆中搜索相关命令, 从而简化了用户对系统施加命令的过程.

菜单除了展示命令外, 还把一个应用内的所有命令通过某种易于用户理解的方式 (如层级方式) 组织起来, 以便于用户查找和遍历. 但当用户熟悉了一个菜单组织体系后, 会对其他不同组织方式有所抵触. 例如微软 Office 2007 的用户界面使用了功能上更为清晰、结构上更为简单的菜单组织设计, 可以帮助用户提高工作效率^[13], 但因为该设计和用户已经熟悉并使用了很久的菜单形式差别很大, 需要用户重新学习, 所以曾被用户所诟病.

3.1.3 点击设备

点击设备的作用是定义用户操作的方式. 现在常用操作方式是使用诸如鼠标、轨迹球、触控板等点击设备去操控界面中光标的二维平面位置. 如果一个交互系统可以直接感知用户的操作 (如触摸屏、笔交互), 用户则通过手指或笔来直接指明交互位置, 而无需借助点击设备控制的光标.

3.2 WIMP 范式与 Norman 的交互鸿沟问题

在人机交互领域, Norman 的交互鸿沟问题^[14] 指明了界面设计中的几个主要挑战, 并极大地影响了界面设计的研究和实践. Norman 所指的鸿沟包括两方面: 执行鸿沟和评估鸿沟. 执行鸿沟是指用户在面对一个界面时, 缺乏足够的信息来决定如何对系统施加命令; 而评估鸿沟则指用户在施加了一个命令后, 缺乏足够的信息来判断系统是否对该命令做出了正确的响应.

WIMP 界面范式从某种程度上讲, 对这两个鸿沟都给予了解答. 图标和菜单明确给出了用户可以执行的命令, 用户需要做的就是去找到相关的命令; 点击设备则帮助用户对一个有限的命令集进行搜索, 并在找到所需要的命令后点击并执行该命令. 窗口则部分解决了评估鸿沟的问题, 因为用户可以比较容易确定去哪里寻找需要评估的信息. 当然, 评估鸿沟的解决还主要依赖诸如任务进度条、任务结果对话框等信息反馈部件.

这里图标、菜单和基于点击设备的点击行为共同解决了交互界面中功能预示性 (affordance) 的问题. 功能预示性^[15] 是指一个交互部件能让用户直接感受到该部件使用的方式. 图标在设计上往往以一个按键的方式出现, 表明这个图标是可以按下的. 菜单作为一个命令集的组织方式, 具有可被下拉、被遍历的模式, 用户可以通过光标和点击来寻找相关的命令.

3.3 WIMP 界面范式与桌面 (desktop) 界面隐喻

界面设计中常用到的一个概念是隐喻 (metaphor). 例如人们常提起的桌面 (desktop) 系统这一概念中, 桌面就是一个隐喻. WIMP 和桌面常被同时提起, 但二者在界面设计中有不同的意义.

3.3.1 隐喻的概念

隐喻的本意在语言学领域内是指词义的转移^[16], 后来被扩展为一种认知工具^[17,18], 来指代人们利用已有的经验认识新事物的一种途径和方法.

在人机交互领域, 界面隐喻是用户界面设计的重要组成部分, 它利用用户所熟知的真实世界中的概念和事物来表征界面中抽象的对象和功能. 界面隐喻往往包含了一个界面应该支持的主要任务, 以及任务的执行方式.

界面隐喻的选择需要有多种考虑^[19]. 但界面隐喻的主要目的是提供一个可以共享的心智模型 (mental model), 以帮助研究人员、设计人员、开发人员和用户建立一个关于用户界面的统一认知模型^[20]. 例如桌面隐喻把日常生活中的桌面这一概念引入界面, 让界面设计人员可以根据人们的日常工作行为来设计需要的部件和任务, 同时也让用户把日常生活中的行为迁移到与界面的交互中.

虽然桌面界面隐喻在很长的时间内主导了界面设计, 在界面设计中还有其他不同的隐喻, 例如游戏和虚拟现实中的空间隐喻 (spatial metaphor)^[21]. 该隐喻包含了很多和空间相关的因素, 如用户在空间内的移动、对物体空间位置的控制等.

3.3.2 隐喻和界面范式的关系

界面隐喻和界面范式密不可分却又彼此不同. 界面隐喻侧重于关于用户界面的统一的心智模型, 而界面范式则描述设计一个界面时应该考虑的主要界面组件、应该支持的交互方式等内容^[22].

桌面隐喻通过办公桌面这一为大家做熟悉的概念, 暗示了一个基于桌面的界面应该支持的目标和任务. 如处理的主要目标对象包括桌面上的文件, 以及其他桌面上的通讯录等其他常见对象, 主要任务包括针对这些对象的相关管理任务, 以及在不同任务间的切换. 基于桌面隐喻的交互范式则应该支持这个隐喻所隐含的交互目标和交互任务. WIMP 范式界面下的多窗口设计很适合支持多任务.

需要指出的是, 在桌面这个隐喻下的第 1 个用户界面并不是基于 WIMP 范式的, 虽然 WIMP 范式对桌面任务的良好支持使得人们经常把 WIMP 和桌面相提并论. 最早使用桌面这一隐喻的是 Commodore 64 系统下的 Magic Desk I 用户界面¹⁾, 该界面并没有包含 WIMP 范式下的所有部件.

3.3.3 隐喻和任务、系统使用背景相关

隐喻虽然可以帮助界面设计, 但随着计算机的发展, 用户使用计算机执行的任务可能会越来越复杂, 会超出一个隐喻所能表征的任务. 此时, 这个隐喻反而可能对界面设计形成了制约. 比如桌面隐喻针对的是办公室内诸如文档处理等办公类任务, 但当计算机的功用转向游戏、娱乐等非办公任务时, 桌面这一隐喻就有局限性了.

对于不同的应用, 或许应该采用不同隐喻. 例如包含空间的游戏类信息, 使用空间隐喻更为合适, 而在空间隐喻下, 交互的方式就有所不同, 比如信息的展示不应在一个二维平面上, 交互方式则可能需要能支持三维动作.

3.4 一个界面范式包含的基本要素

通过以上对 WIMP 界面范式的分析, 可以总结出以下界面范式应该包括的基本要素.

3.4.1 明确与交互相关的信息的展示方式

一个界面必须以适当的方式把与交互相关的信息展示给用户. WIMP 范式是以视觉信息为基础, 使用了窗口作为信息展示的主要手段, 而且提供了多方面的信息, 包括和评估任务执行结果相关的信息以及和执行任务相关的信息 (如命令集).

1) <http://toastytech.com/guis/magdesk.html>.

在无法提供视觉信息或者不能占用视觉通道的情况下, 界面设计可以需要考虑其他的信息展示方式, 如音频信息、触觉信息等.

3.4.2 定义具有良好功能预示性的交互对象

一个界面必须具有可以被用户操作的交互对象, 而这些交互对象应该具有良好的功能预示性, 以使用户能容易理解这些对象的功能和使用方式. 在 WIMP 范式下, 图标和菜单展示了可以帮助用户搜索相关命令的视觉信息, 同时也提示用户该如何对其进行操作.

对于不依靠视觉信息的界面, 则需要通过其他感知通道帮助用户了解可交互的对象.

3.4.3 提供直观、易用的交互方式

一个界面需要一个明确的、适当的交互方式, 以便于用户把一个交互任务中的心理目标转化为一个可操作的物理目标. WIMP 范式下的设计往往采用直接操控 (direct manipulation) 的方式^[23], 让用户把一个点击设备的移动和按键行为映射为用户界面中光标的移动和点击. 而这种直接操控的交互方式也可以随设备的技术特点而有所变化, 例如具有可触控显示设备的系统可以使用比鼠标和光标更为直接的手指触摸或笔输入方式.

3.5 Post-WIMP 界面范式

WIMP 界面范式对图形用户界面的成功起了很大的作用, 但随着计算机应用领域的不断扩展, WIMP 范式也在某些方面表现出不足, 例如 WIMP 范式对于面向三维空间交互的界面设计缺乏有效的指导作用^[24]. 因此诸如基于现实的交互 (reality-based interaction)^[25] 和 PGIS^[1] 等 Post-WIMP 界面范式相继被提出. 但这些范式也需要考虑交互过程中信息的展示、交互对象, 以及交互方式等基本问题.

例如针对笔式的 PGIS 界面范式^[10], 使用了纸笔隐喻, 采用纸张作为承载信息体, 把扩充了的笔式交互界面组件和具有标识性质的图标作为交互对象, 而交互方式则是基于笔手势.

4 智能系统对人机交互范式的影响

基于上述的分析, 本节讨论智能系统对人机交互范式的影响. 首先简单讨论智能系统对宏观层面的人机交互研究范式的影响, 然后集中讨论智能系统对界面范式的影响.

4.1 智能系统对人机交互研究范式的影响

随着计算机与人们生活和工作的高度融合, 计算机的应用已经不再局限于早期的计算和文档处理等办公任务, 人与计算机交互的目的往往是为了与他人更好的交流 (如社交网络、协同工作等领域)、与数据更有效的交流 (如可视分析、情报分析等领域), 以及与周边空间和环境更有效的交流 (如智慧家庭、智能教室等领域) 等更为复杂的目的. 目前的智能系统大都还只是改善、协助这些已有的交流目的, 而诸如电影《她》等科幻作品所描绘的那种人与智能系统深度交流的新交流模式还有待时日.

从人机交互领域的 3 个基本范式角度看, 目前智能系统还没有对当前人机交互领域主流的 3 个范式 (尤其是第 2, 3 个范式) 造成直接的冲击. 智能系统只是增强了系统的功能, 虽然极大地改进了人与系统的交互手段, 但还没有对这个领域的基本研究方向和研究方法造成根本性的改变.

4.2 智能系统对界面范式的影响

但智能系统对界面范式的影响很大,表现在基于新型的、自然交互模式的系统在人们生活和工作中的比重日益增强. 本小节对目前常见的智能系统的交互模式进行分析和讨论.

4.2.1 智能系统常见的交互方式

近期出现的一些智能系统,不再依赖基于 WIMP 的交互方式,多采用更为自然的交互方式.

基于语言的交互. 基于语言交互的系统目前最为常见,代表系统包括 Amazo 的 Echo 系列、Google 的 Home, 以及联想的智能音箱等. 该类系统以一个音箱的形式出现,通过麦克风阵列来接收用户的语音指令,借助云计算来识别用户的语音命令,来完成诸如信息检索、日程安排、播放音乐等任务.

与传统交互界面相比,这类语音交互极大地简化了用户在完成一个任务时从心理目标到物理操作的转化过程. 用户可以直接将一个任务的心理目标转化为语音命令这一物理行为,而无需经过与一个图形用户界面交互的一系列物理操作. 例如需要实现买一本书这一心理目标时,在传统 WIMP 界面下,用户要执行访问网站、搜索书籍、将书籍放入购物筐、完成支付等一系列任务,而完成这些任务涉及控制键盘和鼠标(或触摸屏)等物理操作. 与之相比,在语音交互界面下,用户只需要将目的(买某本书)通过语音命令告知系统,而其他的中间任务由系统自动执行,无需用户任何参与. 在这种交互过程中,用户只需要熟悉相关命令即可.

基于身体姿态的交互. 基于身体姿态的交互界面主要用于游戏机和电视等家庭娱乐设备. 这类系统一般通过摄像头等视频设备来捕捉用户的身体运动状态,然后通过算法来识别用户的姿态,当用户的姿态与系统内部预存的命令姿态相匹配时,系统就执行相关的命令(如移动视点、选择物体等). 身体姿态可以是全身的、上肢的、或者只是手部的动作. 这类系统的姿态识别算法具有一定的智能性,可以从个人用户的自然姿态中识别出命令姿态.

人的身体姿态往往是三维的,因此基于身体姿态的交互可以为与空间相关的系统(如三维游戏)提供更好的用户体验. 虽然这类系统还不像基于语音交互的系统普及,但随着传感设备的进步和计算技术的发展,当诸如智能教室^[26]、智能家庭^[27]等人与空间交互的智能系统进一步普及时,对基于身体姿态交互的需求会有所增强.

姿态交互也可以应用于虚拟现实和增强现实系统. 虚拟现实系统往往需要将用户借助数据手套、手柄等辅助设备输入的姿态信息以某种可见的形式呈现在虚拟场景中(如数据手套的状态被渲染为一个虚拟手),以使用户能即时察觉自己的交互行为. 而增强现实则允许用户透过可视屏幕直接看到自己的操作行为(如手的行为). 例如 Microsoft 的 Hololens 和 Kinect 结合起来可以支持自然姿态交互^[28].

面向智能机器人的交互. 目前也出现了一些其他形式的智能系统,如家庭智能机器人. 这类智能系统除了接受语音、身体姿态指令外,还配有其他类型的物理传感器. 进化者公司的小胖智能机器人²⁾就配有距离传感器. 如果能有效地利用这类信息,人与智能系统的交互会更为丰富. 例如系统可以根据智能机器人和用户的距离,将诸如社交距离等微妙但重要的社会因素^[29]纳入交互设计中,使智能系统更具人性化. 但目前融合了这方面信息的智能系统还很少见.

4.2.2 智能系统交互界面的特征分析

本小节根据对 WIMP 界面范式的分析总结出的界面基本要素,从信息载体、交互对象、交互方式等几个方面来分析目前智能系统界面设计的特点.

2) <http://www.efrobot.com/>.

信息的承载. 智能系统的信息承载能力与传统计算机系统相比有较大的变化. 首先很多智能系统所能提供的交互信息十分有限, 往往是简单的反馈信息, 如表示命令接收到、系统正在运算等信息. 而这些信息一般是以类别信息出现, 如不同颜色的显示灯. 这种信息表达的方式在信息容量和质量上都无法和传统计算机系统通过高分辨率显示器上所展示的信息相比.

另外一个巨大变化体现在对用户所需要的关键信息 (非系统状态信息) 的展现方式上. 基于语音的交互系统往往以语音的方式来展现结果, 要求用户利用听觉通道来获取信息. 由于听觉信息不能在空气媒体中常驻, 用户必须记住相关的信息. 人的有限认知能力决定了所展示的听觉信息不能太长、太复杂. 考虑到人类对视觉信息的处理能力 (如记忆容量、记忆的持久性、记忆内容的细致程度) 要优于对听觉信息的处理^[30], 这类基于听觉信息展示的系统 and 基于视觉信息展示的传统计算机系统相比, 在支持需要对复杂信息进行评估的交互任务 (如意图不是很明确的搜索、对非陈述性知识的搜索方面, 会有一些的难度. 例如美国 Activate 公司的一个针对使用 Amazon 和 Google 等品牌的智能音箱用户的使用调查表明, 对智能音箱的使用还局限在播放音乐、查询 (如天气) 等简单功能上³⁾.

交互对象. 在智能系统中, 交互对象往往变成了物理层面的实体设备 (如 Amazon Echo 的音箱), 而不再是界面中的交互部件 (如 WIMP 界面中的图标、菜单等). 用户直接与物理层面的对象进行交互有利也有弊. 从积极的角度看, 用户只需要学习一个层面的交互. 但消极的一方面是因为物理层面的设计往往被纳入产品设计的范畴, 而缺乏设计标准, 因此不同厂家对同一类产品的设计可能相差甚远, 用户需要付出更多的精力和认知资源去学习相关产品. 例如在语音交互下, 不同系统可能设计了不同的命令关键词, 用户可能需要学习这些不同的关键词. 同样的情况也适用于身体姿态交互.

用户通过自然交互方式与这类交互对象交互时的另一个挑战是如何准确判定针对这些对象的有意义的交互行为 (如命令语音、命令手势), 而不是其他无意义的非交互行为 (如日常的交谈、无意识的手势). 目前基于语音的系统主要通过一个关键词来表明有意识的语音命令. 例如使用 Amazon 的 Echo 系统时, 用户需要用 “Alexa” 这一关键词来呼叫系统, 然后施加命令. 对于可以近距离接触的设备, 这种呼叫命令也可以通过其他模式, 比如触摸 (如进化者公司的小胖机器人) 和特定的按键 (如手机的实体物理键来进行语音交互).

交互方式. 智能系统对用户影响最大的是交互方式. 在智能系统下, 用户无需借助键盘和鼠标等交互工具, 可以通过语言和姿态等自然行为直接与系统交互.

如何能正确理解用户这些自然交互行为的真实意图对智能系统来说还是一个挑战. 当用户实施这类行为时, 很多情况下, 用户的心理目的也不是很明确. 在 WIMP 范式下, 用户是通过与系统的反复交流而达到其目标的, 这个交流过程通过执行一系列的命令让意图更清晰、结果更准确.

当前的智能系统可以通过某些方式来了解用户的意图, 如查询用户以前的交互历史或与其他大众用户的相关行为做比对, 但大部分系统还缺乏实时准确判断用户意图的能力. 基于 Grice 准则^[31], 人与人之间有效的交流要求双方所交流的信息是充分必要的、真实有据的、贴切议题的, 并在交流的方式上双方能做到避免晦涩和歧义、达到简练而有条理. 电影《她》中的智能系统展现出了完美和交流能力, 达到了这些要求, 但真正实现这样的智能系统还有很长的路要走.

更为自然的交互方式也可能带来用户认知负担方面的挑战. 智能系统需要用户记住系统的相关命令, 因此用户在交互中的认知任务又回到了回忆 (recall) 任务这一层面. 虽然和命令行范式下的回忆任务相比, 智能系统中对命令的回忆可以不必很精确, 但记住并准确回忆智能系统中的众多命令 (如 Amazon 的 Echo 有 20 大类, 几百条命令), 对用户来说还是一个不小的挑战. 因此, 智能系统应该有工

3) <https://www.prnewswire.com/news-releases/activate-releases-key-insights-for-the-future-of-tech-media-in-2018-300538862.html>.

帮助用户快速有效地搜索可能的命令。

智能系统中的自然交互方式所带来的另外一个问题是对命令输入的信息质量的要求。目前的智能系统大多依赖统计的方式来提取用户行为中可以 and 系统命令相匹配的典型行为, 但当用户的交互行为不够典型时, 就会导致系统的错误响应或无响应。例如一个用户的口音可能会影响基于语音交互系统的功能⁴⁾, 而对于基于身体姿态交互的系统来说, 用户的固有习惯姿态与系统期望的姿态的不同也是一个挑战。

5 RMCP 界面范式

随着智能系统交互方式、交互对象等因素的改变, 界面设计范式也需要随之而变。在上述分析的基础上, 本节提出一个针对智能系统用户界面设计, 包含了角色 (role)、交互模态 (modal)、交互命令 (commands)、信息呈现方式 (presentation style) 等部件的界面范式 RMCP。

5.1 角色

角色在这里指一个智能系统所能实现的功能。一个智能系统的角色可以根据日常生活中能完成类似系统功能的人的角色来确定。目前的智能系统常常被描述为智能助理 (如 Amazon Echo)、智能玩伴 (如小胖机器人)、智能着装设计师 (如 Amazon 的 Echo Look⁵⁾)、智能保姆 (如美泰的智能保姆⁶⁾) 等。

角色这个部件在 RMCP 范式中的作用是提供智能交互的界面隐喻。目前个人计算机系统是一个通用系统, 可以完成多种任务, 如办公、游戏、社交、娱乐、信息搜索等。而智能系统则往往专注一个或一类特定的功能。角色这个部件把界面隐喻显性化, 可以帮助研究人员、设计人员、开发人员、用户对一个智能系统建立一个清晰的心智模型, 清楚地理解该系统的功能。例如智能助理中的助理这个角色表明系统可以作为一个助手来协助工作, 而智能玩伴中的玩伴角色则表达了一个可以一起娱乐、学习的伙伴这一身份。

一个界面所选择的角色应该能恰当地反映系统的功能。一个与系统功能不符的角色可能会导致错误的用户心智模型, 让用户对系统功能和交互能力抱有不切实际的期待, 从而造成使用问题。例如助理和玩伴这些角色都能够在一定程度上反映系统的一个功能, 但也应该认识到在现实生活中, 一个真实的助理对其所服务的对象的工作习惯有着深入的了解, 可以通过服务对象的语言和行为习惯等多方面的信息来获取其真实意图。同样一个真实的玩伴也有能力去察觉同伴的语言和行为动作本身以及之外的含义。而目前的智能助理和智能玩伴还仅仅局限于某一特定的交流通道 (如仅仅是语言或动作)。如果用户期望他们的智能助理或智能玩伴不仅能做到理解语言命令, 还能对交互时物理环境、社会环境等因素做出正确的判断和反应, 那他们必定会对现有产品感到失望。

5.2 交互模态

虽然角色往往暗示了可能的交互方式, 一个智能系统的界面设计还需要明确其所支持的交互模态。例如智能助理这个角色隐含了诸如语言、表情、动作等现实生活当中人与人之间自然交流模式, 但

4) <https://www.wired.com/2017/03/voice-is-the-next-big-platform-unless-you-have-an-accent/>.

5) <https://arstechnica.com/gadgets/2017/04/amazon-latest-echo-product-is-a-hands-free-fashion-camera-and-virtual-stylist/>.

6) <https://www.engadget.com/2017/01/03/mattel-aristotle-echo-speaker-kids/>.

不是每个智能助理系统都能做到对这些交互模态的全方位支持. 一个界面设计应该明确用户可以采用哪种、或哪几种交互模态.

丰富的多通道交互可以帮助系统更准确地理解用户的意图, 更好地了解交互过程中的背景信息^[32]. 但目前的智能系统还只是侧重于单一交互模态. 虽然有些产品融合了多个输入通道, 但每个输入通道的信息往往针对独立的任务, 并没有被综合起来以更好地理解用户的意图. 例如小胖机器人具有语音输入和触摸输入等多个通道, 但这触摸通道只是用于激活系统等系统任务, 没有被用来协助对用户语音命令的理解.

5.3 交互命令

一个界面应该根据交互模态确定相应的交互命令. 对于语音交互模态, 交互命令包括能被识别的关键词; 对于身体姿态交互模态, 交互命令则为有意义的动作和姿态. 随着传感技术和算法的发展, 越来越多的交互模态可能会出现在智能系统中, 这些模态都需要其相应的交互命令.

5.4 信息呈现方式

信息展示的方式可以通过多个感知通道 (如听觉、视觉、触觉等). 一个系统可以根据其功能、信息的类别、用户的具体交互环境等多种因素来确定选择哪个, 或哪几个感知通道来展示信息. 例如对于数量较大、复杂度较高的信息, 视觉通道可能更为有效.

目前的智能系统的信息呈现方式还比较初级, 以确认输入信息和展示系统结果为主. 在确认输入信息方面, 目前的设计往往假设用户的命令是确定的、清晰的, 但当用户采取语言和姿态等自然交互方式与系统交互时, 语言和姿态的歧义性和多样性意味着用户提供的命令具有不确定性, 因此智能系统可能需要新的信息反馈确认机制来降低这种不确定性. 在现实世界的人与人的交流中, 我们通过多个通道的信息来确认交流信息, 如语言、动作等主要信息通道, 以及眼神、表情等能辅助信息通道. 智能系统也可以考虑采用多通道方式, 帮助用户和系统能更全面地了解对方.

系统结果的展示方式取决于一个系统的功用和所依赖的技术. 智能音箱类系统使用的音频展示方式可以很好地服务于对简单的、陈述性信息的展示. 但对于复杂的信息, 可能就需要视觉通道, 例如 Amazon 的 Echo Show⁷⁾ 在传统的 Echo 系统的基础上增加了一个显示屏用于视觉信息的展示 (以及触觉输入).

5.5 RMCP 的应用

这一小节通过一个假想的智能系统——家庭智能护理机器人——来说明 RMCP 的应用. 随着人工智能和机器人技术的快速发展, 智能护理系统的概念开始引起人们的注意⁸⁾⁹⁾. 设想这么一个面向家庭的智能护理机器人: 该机器人可以在饮食、起居等方面帮助老人、病人等有特殊需求的用户. 这个智能机器人的交互设计可以通过 RMCP 范式来进行.

7) <https://gizmodo.com/the-amazon-echo-show-is-the-best-dumb-smart-machine-in-1796380588>.

8) <https://www.usnews.com/news/healthcare-of-tomorrow/articles/2017-10-31/hospitals-utilize-artificial-intelligence-to-treat-patients>.

9) <https://money.usnews.com/money/personal-finance/family-finance/articles/2017-03-31/how-artificial-intelligence-could-change-the-face-of-senior-health-care>.

5.5.1 角色

根据这个系统的功能, 其角色可以被定义为家庭护工或家庭护士. 在这个角色的基础上, 可以总结出用户期待系统所具有的功能, 包括与被服务对象的日常交流、观察被服务对象的行为、检测被服务对象的体征状态、根据被服务对象的需求提供相应的协助和服务等.

5.5.2 交互模态

系统的交互模态可以在其角色和相应功能的基础上给予定义. 护工和被服务对象应该可以直接通过语言进行交流, 因此交互模态应该支持语音通道. 同时, 护工需要观察被服务对象的行为, 所以基于视频的姿态交互也应该是一个重要的交互模态, 而且对于该系统来说, 除了能识别被服务对象的有意识的姿态命令外, 还应该具有能捕捉被服务对象异常行为的能力 (如摔倒). 此外, 该系统还应该具有能获取被服务对象体征信息的能力, 考虑到这类信息的复杂性、专业性和数据量, 用户和系统传递此类信息的交互通道应该具有比语音和姿态通道更高可靠性和更高带宽的通道, 因此被服务对象可以通过无线数据通道把相关信息通过便携设备 (如手机) 传递给系统, 或者控制所佩戴的医疗器械让其与智能系统直接进行数据交换.

5.5.3 交互命令

交互命令可以根据系统的各类功能和所需的交互通道来确立. 智能护理机器人的交互命令应该包括相关的语音、身体姿态等由用户发出的主动交互命令, 同时, 该系统还应该把摔倒、抽搐等异常身体行为作为非主动交互命令. 在体征信息传递等环节中, 交互命令则可以考虑传统的系统信息交换方式 (如客户机服务器模式).

5.5.4 信息呈现方式

信息呈现的方式需要考虑声音、图像、表情, 以及身体动作等多种方式. 语音信息可以及时对用户的命令做出反馈, 同时系统的语音音调也可以根据用户的需求而变化, 以增强用户对系统的信任感. 当用户需要关于健康、保养等方面的信息时, 智能护理机器人可以通过视频来为用户展现复杂、生动的信息. 如果该机器人可以控制面部表情和身体动作, 那么面部表情和肢体动作也可以用来丰富系统信息呈现的方式, 让系统和用户的交互更为融洽.

6 总结

随着智能系统的快速发展和与人们生活的深度融合, 如何设计有效、易用的智能系统交互界面成为当前的一个重要课题. 在人机交互的发展过程中, 界面范式为有效的界面设计奠定了基础. 但在智能系统交互设计方面, 目前还缺乏能指导界面设计的界面范式.

本文力图探索能支持智能系统界面设计的界面范式. 通过对范式这一概念的起源及其意义的追溯, 本文介绍了人机交互领域对范式的不同理解, 讨论了 WIMP 界面交互范式的主要因素. 本文还分析了当前常见的智能交互界面的特点. 分析侧重于基于语音和身体姿态交互的界面, 并讨论了这些交互系统的优点以及潜在的问题.

在上述分析的基础上, 本文提出了一个针对智能系统的功能角色、交互模态、交互命令, 以及信息呈现方式的 RMCP 界面范式. 在系统的功能角色方面, 该范式强调考虑一个系统以什么样的身份出

现来帮助人们的生活和工作. 在交互模态和交互命令方面, 该范式建议系统应考虑自然交互设计, 力求降低用户在学习和实施交互行为过程中需要付出的认知资源. 在信息的展示方面, 该范式则提出在设计中应该根据信息的特点、交互的场合等因素, 选择相应的通道来为用户提供所需的各种交互信息.

参考文献

- 1 Shi L, Deng C Z, Dai G Z. Implementation of PGIS: a type of Post-WIMP user interface. *J Image Graph*, 2010, 15: 985–992 [石磊, 邓昌智, 戴国忠. 一种 Post-WIMP 界面: PGIS 的实现. *中国图象图形学报*, 2010, 15: 985–992]
- 2 El-Assady M, Sevastjanova R, Sperrle F, et al. Progressive learning of topic modeling parameters: a visual analytics framework. *IEEE Trans Visual Comput Graph*, 2017, 24: 382–391
- 3 Kuhn T S. *The Structure of Scientific Revolutions*. 1st ed. Chicago: University of Chicago Press, 1962
- 4 Kuhn T S. *The Structure of Scientific Revolutions*. In: Jin W L, Hu X H, eds. Beijing: Peking University Press, 2004 [Kuhn T S. 科学革命的结构. 金吾伦, 胡新和译. 北京: 北京大学出版社, 2004]
- 5 Gray J. eScience: a transformed scientific method. In: *The fourth Paradigm: Data-Intensive Scientific Discovery*. 2009. <https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/>
- 6 Harrison S, Sengers P, Tatar D. Making epistemological trouble: third-paradigm HCI as successor science. *Interact Comput*, 2011, 23: 385–392
- 7 Card S, Moran T, Newell A. *The Psychology of Human-Computer Interaction*. Mahwah: Lawrence Erlbaum, Associates, 1983
- 8 Carroll J M. Human-computer interaction: psychology as a science of design. *Annu Rev Psychol*, 1997, 48: 61–83
- 9 Dong S H, Chen M, Luo J, et al. The model, method and instance of multimodal user interface. *Acta Sci Nat Univ Pekinensis*, 1998, 34: 231–239 [董士海, 陈敏, 罗军, 等. 多通道用户界面的模型, 方法及实例. *北京大学学报 (自然科学版)*, 1998, 34: 231–239]
- 10 Dai G Z, Tian F. *Pen-based User Interface*. Hefei: China Science and Technology University Press, 2014 [戴国忠, 田丰. 笔式用户界面. 合肥: 中国科学技术大学出版社, 2014]
- 11 Green M, Jacob R. SIGGRAPH'90 workshop report: software architectures and metaphors for non-WIMP user interfaces. *ACM SIGGRAPH Comput Graph*, 1991, 25: 229–235
- 12 Nielsen J. Noncommand user interfaces. *Commun ACM*, 1993, 36: 83–99
- 13 James J. Word 2007: lessons on usability. <http://www.techrepublic.com/article/word-2007-lessons-on-usability/>
- 14 Norman D A. *The Psychology of Everyday Things*. New York: Basic Books, 1998
- 15 Gibson J J. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979
- 16 Liu Z X. Interpretation of the theory of metaphor. *J Anhui Univ Technol (Soc Sci)*, 2007, 24: 126–127 [刘正霞. 隐喻理论解读. *安徽工业大学学报 (社会科学版)*, 2007, 24: 126–127]
- 17 Wang M. An exploratory study of the understanding of user interface metaphors: the case of metaphoric icons. Dissertation for Master Degree. Wuhan: Central China University of Science and Technology, 2011 [王敏. 界面隐喻理解的探索性研究 —— 以隐喻性图标为例. 硕士学位论文. 武汉: 华中科技大学, 2011]
- 18 Lakoff G, Johnson M. *Metaphors We Live by*. Chicago: University of Chicago Press, 1981
- 19 Carroll J M. Human computer interaction — brief intro. In: *the Encyclopedia of Human-Computer Interaction*. 2nd ed. <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed>
- 20 Szabó K. Metaphors and the user interface. 1995. <http://www.katalinszabo.com/metaphor.htm>
- 21 Ark W, Dryer D, Selker T, et al. Representation matters: the effect of 3D objects and a spatial metaphor in a graphical user interface. In: *People and Computers XIII*. Longdon: Springer, 1998
- 22 Lyu F, Tian F. *Reality-Based Interaction: Method and Practice*. Beijing: Electronic Industry Press, 2017 [吕菲, 田丰. 基于现实的交互界面: 方法和实践. 北京: 电子工业出版社, 2017]
- 23 Shneiderman B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 2nd ed. Boston: Addison-Wesley, 1992
- 24 Dam A V. Post-wimp user interfaces. *Commun ACM*, 1997, 40: 63–67
- 25 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 ACM SIGCHI Conference on Human Factors in Computing Systems*, Florence, 2008. 201–210
- 26 Shi Y, Xie W, Xu G, et al. The smart classroom: merging technologies for seamless tele-education. *IEEE Pervas Comput*, 2003, 2: 47–55
- 27 Aldrich F K. Smart homes: past, present and future. In: *Inside the Smart Home*. London: Springer, 2003, 17–39

- 28 Davies C, White J, McAllister A, et al. A toolkit for building collaborative immersive multi-surface applications. In: Proceedings of the 2016 ACM on Interactive Surfaces and Spaces (ISS'16), Niagara Falls, 2016. 485–488
- 29 Hall E T. The Hidden Dimension. 2nd ed. New York: Anchor Press, 1990
- 30 Cohen M, Horowitz T S, Wolfe J M. Auditory recognition memory is inferior to visual recognition memory. Proc Natl Acad Sci, 2009, 106: 6008–6010
- 31 Grice H P. Logic and conversation. In: Syntax and Semantics III: Speech Acts. New York: Academic Press, 1975
- 32 Dumas B, Lalanne D, Oviatt S. Multimodal interfaces: a survey of principles, models and frameworks. In: Human Machine Interaction. Berlin: Springer, 2009

Interaction paradigm in intelligent systems

Xiaolong ZHANG^{1,2*}, Fei LYU³ & Shiwei CHENG⁴

1. College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030027, China;
 2. College of Information Sciences and Technology, Pennsylvania State University, University Park 16801, USA;
 3. School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications, Beijing 100876, China;
 4. School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China
- * Corresponding author. E-mail: zxl_psu@foxmail.com

Abstract This paper discusses the user interface design paradigm for intelligent systems. We first introduce the concepts of paradigm in science and human-computer interaction (HCI). After discussing the WIMP (windows, icons, menus, a pointing device) paradigm, the most influential user interface design paradigm in HCI, we summarize the major components a user interface paradigm usually takes into consideration, and based on this summarization, we analyze the features as well as limitations of the user interfaces of some intelligent systems currently available on the market. We also propose a user interface design paradigm, RMCP (role, modal, commands, and presentation style), for intelligent systems, which emphasizes the consideration of the role(s) that an intelligent system plays in a user's life and work, the interaction modal(s) that the system can accommodate, the commands that the system supports, and the presentation style(s) that the system adopts for information outputs.

Keywords human-computer interaction, interface paradigm, intelligent system



Xiaolong ZHANG received his Ph.D. degree in information science from the University of Michigan, Ann Arbor, USA in 2003. Currently, he is an associate professor at the Pennsylvania State University, and a visiting professor at Taiyuan University of Technology. His research interests include human-computer interaction and visual analytics.



Fei LYU received her Ph.D. degree in computer applied technology from the University of Chinese Academy of Sciences in 2013. Currently, she is an assistant professor at Beijing University of Posts and Telecommunications. Her research interests include human-computer interaction and user interface design.



Shiwei CHENG received his Ph.D. degree in computer science and technology from Zhejiang University in 2009. Currently, he is an associate professor at Zhejiang University of Technology. His research interests include human-computer interaction, ubiquitous computing, and user interface design.