

女计算机工作者专栏

做好人机交互研究

史元春
清华大学

关键词：人机交互

人机交互（Human Computer Interaction, HCI）研究人与计算机系统之间自然高效信息交换的原理与技术，实现为由多种模态的输入输出软硬件接口所构成的用户终端界面，形成特定的交互模式。如图1所示，接口分为用户输入数据处理的输入接口和机器处理结果反馈的输出接口。人的交互意图在脑中产生，今天的生命科学和脑电技术尚不能实现直接读脑写脑（图1中表示为虚线），交互意图需要通过外周神经系统下的行为动作表达出来，可以是操控工具，也可以是语音和动作的自然表达，输入接口的主要任务是捕捉和处理人的外在行为；机器处理结果的呈现要符合人的感知认知特点。

人机交互是研究内容较广的领域，今天大量的识别问题，可算作广义的人机交互研究内容，笔者就是在本科阶段，也就是20世纪80年代开始在语音识别、计算机视觉的学习和研究中接触到交互性这个关键词的，这些识别问题因其明确的计算任务，早已成为专门的研究方向。也正因为这种密切关系，

1999年清华大学计算机系面向学科方向进行机构改革（重组基层科研单位，成立了5个研究所）时，基于语音、视觉、图形学、多媒体、普适计算研究的教研力量，成立了人机交互与媒体集成研究所（我于2003~2018年任所长），这也是国内第一个以人机交互命名的科研单位。那时，我负责的863项目中有基于内容图象检索的用户检索请求表达机制，985学科建设项目中智能空间（smart space）人机交互技术等研究内容。除了数据处理，研究闭环中还涉及我之前并不很熟悉的人因建模与测试等内容。基于这些研究，我开始参考和开展更具交叉学科特色的人机交互原理和技术的研究。事实上，理解、学习这个新领域的研究方法，逐渐建立起自己的研究方法并取得硬成果，是一个比较长期的过程。在此过程中，我从2002年起先后给本科生和研究生开设了3门人机交互课程，这也是一个教研相长的过程，系统性、理论与实践相结合的长期教学对科研的启发有着难以替代的重要作用。当时，国内尚无相关方向的学术社区，我在CCF与同事们一起发起创立了普适计算专业委员会（2005年）、人机交互专业委员会（2013年）。学术社区健康发展，越来越多的学术和产业人士关注和开展着这个领域的研发，如今是计算无所不在、交互即用户产品的时代，人机交互更是产品创新的关键技术。

本文首先简要介绍一下人机交互研究的重要性和特点，然后分享笔者在几个研究层面上的研究思路和方法。

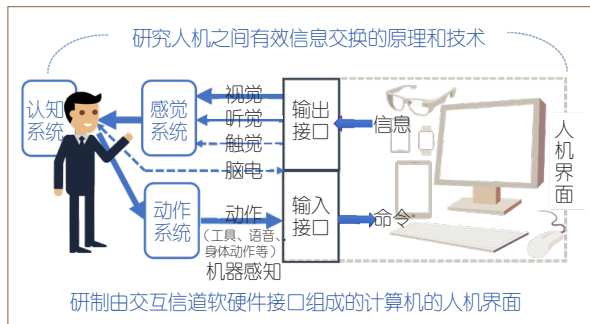


图1 人机交互研究示意图

理解人机交互研究

GUI划时代，图灵有三奖

首先以 ACM 图灵奖为例，简要说明人机交互研究实践的重要性。建立原理、创新技术、构建系统在人机交互研究中都是必要的。

吉姆·格雷（Jim Gray）在 1999 年的图灵奖获奖演说中指出，计算机的研究开发有三个主要的脉络，即巴贝奇问题（如何构造计算机系统）、布什问题（人机之间的关系和如何使用计算机）以及图灵问题（机器智能）。

万内瓦尔·布什（Vannervar Bush）在 1945 年发表的文章“*As We May Think*”中设想了未来的人机（注：当时数字电子计算机尚未问世）关系。1960 年，布什在 MIT 的年轻同事、拥有心理学博士学位的计算机科学家利克莱德（J.C.R. Licklider）将其阐释为人机共生（man-computer symbiosis）。人机共生被认为是人机交互的启蒙观点，不仅在于论文本身的影响力，还在于利克莱德通过美国国家科技计划（在布什的建议下利克莱德从 MIT 到美国国防部高级研究计划署（ARPA）任科技官员）直接领导和支持了系列交互式计算研究项目，研制出了分时操作系统、图形与可视化、虚拟对象操控、互联网络与远程交互等创新技术，奠定了 20 世纪 70 年代产生的划时代的交互模式——图形用户界面（Graphical User Interface, GUI）的技术基础。

GUI 因其界面元素组成又被称为 WIMP（Windows, Icons, Menu, Pointing Device），它成就了 20 世纪 80 年代崛起的 PC 产业，这 3 个英文缩写的技术术语中，分别产生了一位图灵奖获得者，他们是：

• 伊凡·苏泽兰（Ivan Sutherland）：图形学和虚拟现实之父，1963 年提出 Sketchpad: A Man-Machine Graphical Communication System，1988 年获得图灵奖。他的研究使交互界面得以从不可见的抽象转向可见的形象。

• 道格拉斯·恩格尔巴特（Douglas Engelbart）：

鼠标之父，1963 年为其 Augmenting Human Intellect / oNLine System 发明鼠标，1997 年获得图灵奖。鼠标让交互得以从人脑记忆搜索转向动手在显示屏上浏览选择。

• 艾伦·凯（Alan Kay）：PC 和 SmallTalk 之父，1971 年提出现代笔记本电脑原型 Dynabook，2003 年获得图灵奖。他的发明让交互作为操作系统功能得以从形式语言命令转变为所见即所得。

此外，1975 年获得图灵奖的艾伦·纽厄尔（Allen Newell）在对人的感知认知能力和运动控制力建模有重要贡献。

1979 年，年轻的乔布斯拜访施乐 PARC 研究中心时，看到了施乐新发明的 GUI，相比当时的文本命令行界面，施乐 GUI 的程序图标、窗口化、下拉菜单和绚丽的图像效果把乔布斯狠狠震撼了一把。“仿佛蒙在我眼睛上的纱布被揭开了一样。”《乔布斯传》中如此描述乔布斯当时的感受，“我看到了计算机产业的未来”。之后，苹果公司将 GUI 迅速移植到自己的产品上，凭借先进的交互技术，苹果创造了产业奇迹。

扩增发展，喜新不厌旧

计算机的历史只有短短几十年，但它已经并继续深刻地改变和影响社会发展，而在计算机的发展中，HCI 技术是一项引领性的技术，不断突破用户使用计算机的难度瓶颈，新的交互模式极大扩增（数量级的扩大）了用户规模和应用场景。

如图 2 所示，由于交互接口的变革，人机关系发生了重大变化，计算机越来越好，从 20 世纪 80 年代开始，GUI 的出现，带来了所见即所得的图

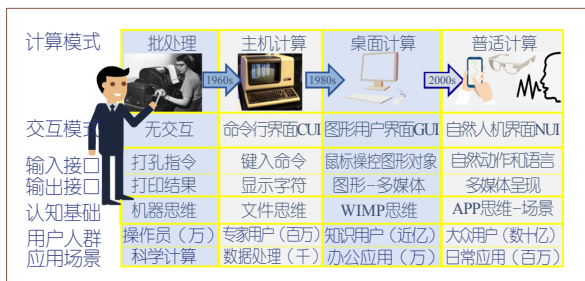


图2 交互模式与计算模式的演化

形化直观呈现和鼠标的直接操控,较之前的字符用户界面(Character User Interface, CUI),也称为命令行界面(Command Line Interface, CLI)的形式语言命令,极大减轻了复杂性、抽象性、记忆负荷重对用户的限制,计算机应用得以普及,走出实验室和机房,走向办公室和家庭;新世纪,计算机用户多半都不曾用过鼠标等交互工具,直接用手与触屏交互——这是抛弃专门的交互设备、离开特设的工作台面后的又一次革命,可以将计算机握在手中、随时可用。正是由于交互模式的变革,计算机得以从卧在机房、立在桌上、握在手中到戴在身上和融入环境,计算无所不在。

今天,随着计算和物联技术的进步,越来越多的设备和器具都具有传感和计算能力,我们已经进入了普适计算的时代,人们和多种多样的设备将发生更多、更频繁也更自然的交互。我们将在多种场景下,通过多种模态的接口,用自然语言、语义丰富的手势,甚至是日常行为对手持设备、家电设备、新型穿戴设备、机器人、交通工具等实现高效的信息访问和获得服务。而用户对于这些终端产品的价值判断,从能用转向好用,在很大程度上取决于人机交互的性能。

需要说明的是,人机交互技术的迭代,并不是简单的更新换代,而是扩展出新的模式,增加了新的应用和人群,而同时“旧”的模式仍然存在和发展。比如,因其高效率,至今,CUI仍是计算机专业人士的首选工作界面,而AI支持的自动补全和容错功能也减轻了用户的记忆和表达负担。这是因为,人机交互研究每有新成果,就是在计算机中新建立了一定的适应人的感知、认知和行为的能力,是对人的能力的不断发现和累积的过程,从人机信息交互的角度看,是带宽增加的过程。

交叉领域,计算融人机

在计算机专业中,人机交互是典型的交叉学科研究领域,从其理论体系到技术实现,都不仅来自单一学科。在理论体系方面,除了信息科学(信息论、计算机科学与技术、计算语言学等)的基础,

HCI最早发端于人机工程学,之后从中独立出来,又接受甚至强调认知心理学、行为学和社会学等某些人文科学的理论指导;在技术实现方面,仅从图1中的交互信道看,传感、信号和数据处理涉及多种技术,与之相关的是多感知、多通道、多模态、多媒体等以“多”统称的技术。

不同的学科在人机交互领域相聚,是为了创造人机之间适应的、自然的、高效的、和谐的相融共生关系,这种关系的建立需要建模、算法所实现的由交互信道软硬件接口组成的计算机的人机界面,而这些都需要通过计算设计实现。如,人机交互领域的一本经典著作是*The Psychology of HCI*(1983年出版,Allen Newell是作者之一),名为心理学,但其中内容都是“可计算的”。又如,费茨1954年在香农信息论基础上提出的人手操控精准设备的运动控制能力模型Fitts' Law(费茨定理)非常简洁:

$$T = a + b \log_2(D/S + 1) \quad (1)$$

其中 a 、 b 是与输入设备(如点选面板、鼠标、触摸板、触屏等)有关的常量, D 为目标对象的大小, S 为从当前位置到目标对象的距离,是变量。公式(1)看似简单,但在实现技术上,由于人的手眼协调性(如图3所示)、手运动的(一般会遵循粗粒度到细粒度的原则移动)控制平面与眼睛看的(视线移动)显示空间的映射关系(称为C/D Ratio)、软件对输入设备信号的处理、软件对界面元素分布的感知等可变因素,建立特定的费茨定理公式,并提供高拟合度的鼠标或触控输入技术、界面布局和动态优化技术,成为决定GUI产品用户体验优劣的关键技术(在同样的交互任务上,不同的技术实现,效率可能有数倍的差异),往往还有专利保护的技术。理解、分析费茨定理及其作用,进行数据采集和回归实验,讨论优化方案,也一直是我所讲授的人机交互课程的重要内容,学生通过实验体会到,好的人机交互技术,其内在算法能很好地与人的运动感知认知能力相融,用户的体验反而是无感的,即所谓的用而不觉。

在这个学科交叉的研究领域,挑战和动力往往来自本来熟悉的计算领域之外,继承和发展计算领

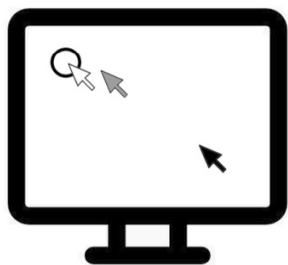


图3 典型手眼协同任务——目标选择：用户将鼠标从右下角初始位置移动到圆形目标位置。鼠标移动过程类似黑色→灰色→白色

域的价值观和方法论才能培养自己发现问题和解决问题的能力。以我们2006年发表在人机交互旗舰会议ACM CHI上的第一篇口头报告论文为例，我们研制大屏幕非接触式交互时，发现直接识别激光笔投射光点不仅有一定的识别鲁棒性问题，还有多用户时物理激光笔的安全性以及人不能抑制的光路延长后的光点抖动造成的点选不准问题。我们直击后者人因造成的难以规避的难题，而没有纠缠前者的技术改进，提出了Direct Pointer以手持摄像头（可嵌入手持设备）识别人手指向所渲染出的屏幕指示光点，通过闭路反馈的自校准机制完成对指示光点的实时（算法在摄像头采样周期内完成）定位和追踪，软光点不仅有激光笔投射屏幕光点的效果，还保证了远距离交互的连续精准和有效范围。由于手持设备本身是计算终端，还可提供多用户ID和丰富的交互原语。最后，通过国际标准的测试方法，与同类问题的其他解决方案（包括摇杆、轨迹球、激光笔等）相比，Direct Pointer在指点设备的输入吞吐量（点选任务的难度指数与完成时间的比值，单位为bps）这一关键性能指标上达到最优。这是CHI录取的首两篇第一完成单位为国内教研机构的正式论文之一。论文投稿前后很长一段时间，我不断反思在绕不开人因的人机交互领域（事实上，CHI的全称就是ACM CHI Conference on Human Factors in Computing Systems），如何明确研究问题的问题域和难点、如何在多约束条件下明确优化的目标、如何设计评测以取得客观性的测试结果，这些问题显然反映的是计算领域形成的价值观和方法论，在交叉领域的研究中被进一步明确和强化了。

做好人机交互研究

目前，计算已经从桌面计算扩展到移动计算，正在进入基于物联网的人机物三元融合时代，人机物之间自然高效的信息交换技术是三元空间基础性的关键技术，是形态各异的三元空间可用性的根本保证。然而，支持用户在非结构空间内的交互行为，面临着多层面的挑战和问题，都需要通过计算建立自然和谐的人机关系来解决，下面结合我的科研实践，简单介绍相应的研究思路和方法。

建模交互能力

人机交互领域的问题、约束和发展空间常常来自对人的交互能力的深入理解，建模交互能力可能是研究中显式开展，亦可能是隐式借鉴的内容。在GUI的产生和发展过程中，《The Psychology of HCI》一书中提出的传统的用户信息处理模型（Human Information Processing Model，简称HIP模型）是基本的交互能力模型。HIP围绕着工作记忆、长时记忆、感知和运动系统解释人的处理信息的过程，但其描述的交互通道范围有限（感知只有视听觉的基本处理，运动只有使用工具的小范围准确操控动作），量化粗略，不能适应当今用户脱离键盘、鼠标甚至是触屏等规范的交互接口，变为手势、姿态、语音等多模态自然交互表达的需求。扩展HIP是发展之道。

视觉注意力（visual attention）是人机交互中的“稀缺资源”，发展无需视觉注意（eyes-free）的技术是近些年探索的热点方向，但多数研究没有从人的动作闭环的底层原理出发优化交互效率，仅仅停留在优化特定交互任务的应用层面。因此我们提出扩展HIP模型，从相对独立的感知（P）、认知（C）、运动（M）三个子系统的处理模型到三个子系统参与动作交互多轮动作闭环处理模型（公式（2）显示了多轮动作闭环的时间开销 T ， k 为任务难度系数， τ 为HIP每个子系统的开销， n 表示交互所需的动作闭环数）。

$$T = n(k_p \cdot \tau_p + k_c \cdot \tau_c + k_m \cdot \tau_m) \quad (2)$$

如图4（a）所示，一个完整的动作闭环会经历

感知、认知和运动三个子系统对应的信息处理阶段,其中运动子系统的信息来源一般依赖于认知子系统,认知子系统的信息来源一般依赖于感知子系统。同时,人体在做出动作回应后,一般会重新回到下一个动作闭环的感知阶段,形成一个完整的闭环。

在此基础上,系统深入地探索通过减少视觉注意力参与环节完成交互动作的方法,如图4(b)、(c),由于优化依据和方向明确,在通用基础动作交互任务上,通过降低视线移动的开销、降低感知周期需要的开销、减少动作闭环的轮数、缩短动作闭环的长度(均可在公式(2)中设定预期更高效的目标)实现的虚拟现实中的自体感知双手之间动作交互、分手软键盘等技术,可提高自然动作交互的效率,显著改善用户体验。如平板电脑分手软键盘输入速度由于省去了手上的主动视觉注意,速度达到iPad的2.6倍。

自然交互强调较低的认识负荷,而传统的认识负荷评测具有较大的主观性,发展客观的评测技术和指标也是交互能力建模的重要方面,是与生理、心理、脑电研究交叉的方向。我们发表在CHI 2015的一篇获奖长文阐释了基于生理参数的认知负荷指标与实时测评技术。

简言之,建模交互能力,发现人的生理感知、认知和行为的自然性机理,本身就预设了解决问题

的方向和目标,是人机交互中的基础性研究。

推理交互意图

自然交互的研究不仅要不断降低记忆和表达负担,在接口上还体现为多种“解放式”(free)的特征:device-free指脱离专门的输入设备(如鼠标、笔)而直接用人体器官作为输入工具,触屏就是以手指为输入工具;eyes-free指无需或减少交互中的视觉注意力;hands-free是解放双手,包括手上不持握设备和不用手交互。在现实空间和虚拟空间中连续、自然的动作交互是正在发展的主流交互方式,但难题是,如何根据模糊的自然行为数据推理人的交互意图?

由于用户行为的观测方式不同(传感器、观测位置),以及针对的应用场景不同,意图推理问题的输入和输出也不同,一般可以将意图推理问题分为分类和回归两类,前者需要根据用户输入的信号,从多个可能的类别中找到对应的分类(如状态检测);后者需要根据交互信号来计算某一具体的数值或指标,以达到提升精度等目的。人机交互新技术研发有一个重要特征是,在技术部署之前是难以收集大量数据的。因此,AI领域很有效的隐变量机器学习类方法由于强烈地依赖于训练数据,适用性很有限。

我们重点研究了能在较小样本中构建特征模型、模型参数一般又具有直接物理意义的贝叶斯方法,提出了基于贝叶斯推理的交互意图推理框架(图5),将任务-情境模型 $P(I|CT)$ 和行为编码模

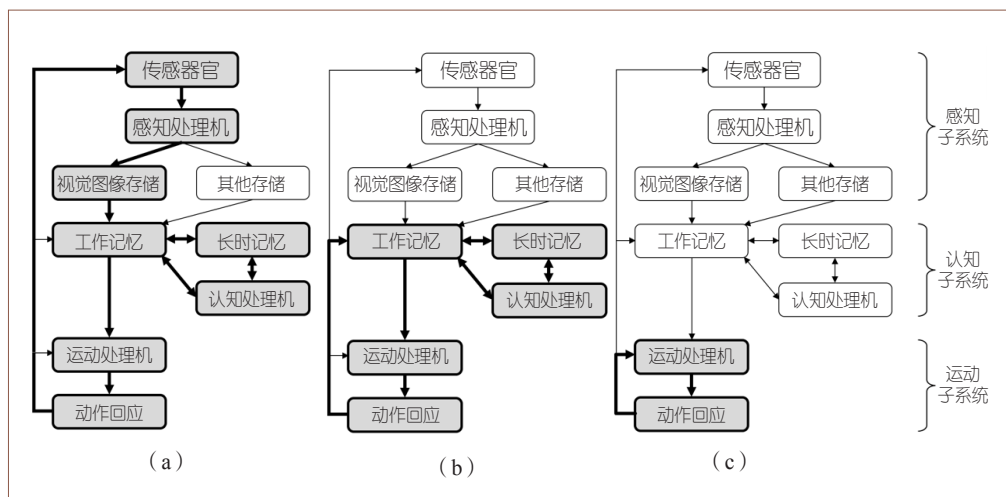


图4 动作闭环HIP模型 (a) 完整动作闭环; (b) 跳过感知子系统的动作; (c) 跳过感知和认知子系统闭环的动作闭环

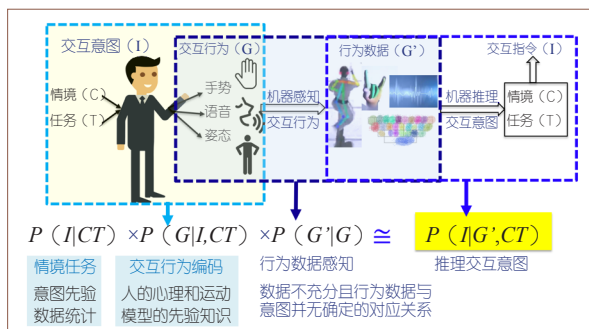


图5 基于贝叶斯的交互意图推理框架

型 $P(G|I,CT)$ 作为先验知识代入贝叶斯推理,可以走出单纯依靠感知数据 $P(G'|G)$ 进行交互语义识别所难以克服的数据不充分和数据-意图难对应的困境,可大幅提高意图识别 $P(I|G',CT)$ 的准确性。与人机交互领域贝叶斯方法的应用相比,机器学习领域的贝叶斯方法的应用主要是在数据中发现知识,而人机交互则是在行为中发现知识。

我们基于该方法为智能手机研制了握持意图识别、软键盘容错输入等国际领先的产品技术,前者有效解决了全面屏高误触难题,后者显著提升了预测纠错能力,提高了输入速度,支持着7亿多用户的日常应用。以文本输入技术为例,由于语言交流是人机交互的一个基本任务,在交互技术发展的历史中,接口形式在变,但这个基本的交互任务一直存在。智能手机普及后,打字从符合人体工学和语言学字符分布特征的物理键盘转变为触屏软键盘,在手机上打字的速度一般是在物理键盘上打字速度的一半,一分钟30个字左右,原因是什么呢?就是人机交互中公认的胖手指难题(fat finger)。我们的手指很难准确点击屏幕键盘这样的精细目标,为了点准,要在视觉注意力参与下手指对准,速度就慢下来了。在基于贝叶斯推理的交互意图推理框架下,交互意图就是打字结果,情境任务模型 $P(I|CT)$ 就是语言模型,可感知的行为数据 $P(G'|G)$ 是用户(非严格对准的打字方式下)在触屏上的点击结果,而作为先验知识的打字行为编码模型 $P(G|I,CT)$ 则是从较小规模用户中采集获得的用户的目标按键和落点实际位置之间的概率关系,语言和行为模型的代入可以在模糊的触摸数据上推理出准确的点击

目标,用户点不准没关系,算法可以猜得准,效果就是软键盘的容错性提高,打字速度也就提高了。这套方法帮助企业在手机、平板、眼镜、大屏幕等设备上的多种软盘大幅提高了性能,在诸如免唤醒语音动作、抽象任务手势、头动、眼动以及健康异常行为等多种新的交互意图推理任务上也很有效,特别是,该方法能有效区分连续活动中人的有意交互动作和无意动作,这在自然交互中非常重要,如,目前的语音交互都需要唤醒词,原因就是不能有效区分有意无意动作,还需要用户明示。其中,建立两类先验知识是关键,即要建立在扎实的人机交互中的情境任务研究和人因研究基础上。

近几年,我指导的两个博士毕业生先后获得中国计算机学会(2018年,易鑫,论文题目《自然文本输入中的贝叶斯推理方法》)、中国电子学会(2020年,阎裕康,论文题目《自然动作输入的“编码-解码”优化方法》)优秀博士学位论文,主要贡献都集中在动作交互意图的推理方法。

优化交互路径

优化在计算机领域是一个常见的术语,交互路径是什么呢?这里先介绍一下交互任务,对于用户而言,交互任务是用户要实现的某个特定的功能,需要在特定的界面上进行多轮的信息交换,可以通过GUI这样的范式界面,也可以是现在的分布物联的多模态接口,这个信息交换过程称为交互路径。完成同一交互任务的交互路径可以是不同的,效率和体验也会是不同的。在约束(各种个性化和场景化)较多的情况下,我们研究提出的交互路径既存在理论最优,实践上也存在多目标联合优化的参考实现。

理论上存在完成交互任务所需的理论最小信息量(基于对交互任务的语义分解计算),即任务信息熵,而交互路径 $(t_1 \cdots t_n)$ 的信息熵可由公式(3)计算获得。对于每步操作提供的信息,不同接口有其对应的计算方式,在GUI上就是公式(1)中的对数部分,手势、语音等模态另有计算方式。

$$H(S) = \sum_{t_1 \cdots t_n \in X^n} p(t_1 \cdots t_n) * \log_2 p(t_1 \cdots t_n) \quad (3)$$

交互路径优化就是以任务信息熵为优化目标,搜索求解并通过路径剪枝和动态生成实现用户输入信息量最小的交互路径。在实践中,通过对界面上用户操作的逆向编译,提取用户完成交互任务过程中表达的交互语义集合。用户的交互语义与具体的交互接口或交互路径无关,仅由交互任务的需要决定。

交互任务及其语义的获得,实践上可以充分利用已经有十多年广泛应用的移动互联网所产生的近千万个应用所携带的交互任务,这就是我们长期开展的界面语义理解工作,称为 NUI Manager,如图 6 所示的下面四层,自动地实现界面元素的分割和基于逻辑层次的自动组织,从无结构的像素信息和移动设备渲染界面的布局信息生成用户理解界面的逻辑结构信息(从 Framework 层直接获取界面信息和交互数据),利用计算机视觉和自然语言处理技术从界面的图像上自动识别出界面控件和操作路径等,并为其增加语义描述。

基于交互语义集合,对交互任务进行拆分和信息量化,根据任务情境、用户特征,采用模态替换(比如盲人只能听和摸、普通人开车时注意力和操作范围的变化等)和互补方法,完成多模态输入交互语义集合的构建和灵活切换,面向用户当前交互任务,实现快捷路径生成和低效路径剪枝,整体上优化用户信息交互效率,如图 6 的上面两层。

人机交互中一个重要的应用领域是信息无障碍, NUI Manager 在无障碍应用中发挥了系统工具的重要作用,为视障人士研发出在智能手机上无须

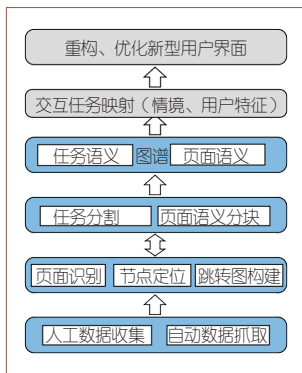


图6 基于交互语义的交互路径优化

改动原应用程序即可转换出更适应其交互能力且能成倍提高交互效率的触觉交互界面、“慧说读屏”等新功能先后上线,近期上线的“拍拍扫码”应用软件更是将广泛应用的各种扫码功能一拍(拍手机上沿的动作)搞定,将繁琐的交互路径缩减为一个

动作,广受用户好评。

普适计算时代,人机关系发生着重重大变革,机器从被动应答者向主动服务者身份转变,交互从单一显式的用户动作表达向隐式机器智能推理与显式用户表达融合的方向发展,在恰当时间与情境下提供用户急需的智能服务成为必然发展趋势。也就是说,人机交互的路径,将从现在用户记忆搜索应用和界面的模式,转换到机器主动感知和推送服务(即图 2 最后一列中,人的认知基础将从 APP 模式转向场景化)、极大缩减交互路径的模式,“拍拍扫码”即为一例。由于用户终端多样,同样功能的服务需要设计实现为每种终端上的应用,开发和优化成本都很大, NUI Manager 有大量的交互应用语义和优化方法,也正被用于重构智能家居、车机等新型终端上的应用。

建交互场景IDE

人机交互实践性强,同时也造成很大的测试和开发成本,开发涉及软硬件的实现,测试验证是必要的研究环节,与一般的计算技术研究不同,人机交互的测试通常需要设计有较多用户参与的控制实验,以获得可信的较为客观的测试结果,这个测试结果往往是发现问题、继续改进技术的基础。近年来,人机交互又呈现出多用户、跨设备的人机物融合交互场景,由于人机物的组成复杂多变,场景定制研发的滞后验证方式往往导致更多轮次开发修正,场景开发面临扩展性差的难题。

我们自 20 世纪 90 年代后期开始研究智能空间系统,在这种始终需要面对的工程难题中,探索并提出多用户混合现实同步交互控制方法,以混合现实方式连接虚拟呈现的场景设备与真实设备,并同步控制不同程度虚实融合的场景中多用户真实交互行为的一致性,在物理设备实现前后均可快速验证互联协议和交互能力。基于该方法,我们研制了人机物交互场景集成开发环境 IDE,称为 NUIX Studio,如图 7 所示,其虚实融合场景中内置大量基础自然交互 I/O 控件、设备控件和场景工具等多层次 API 支持,通过同步交互控制方法将同地或异地的

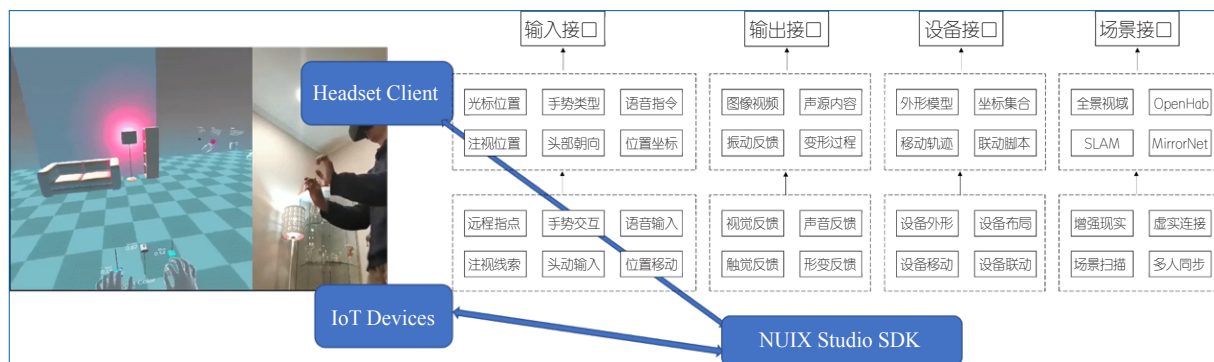


图7 人机物交互场景集成开发环境NUIX Studio

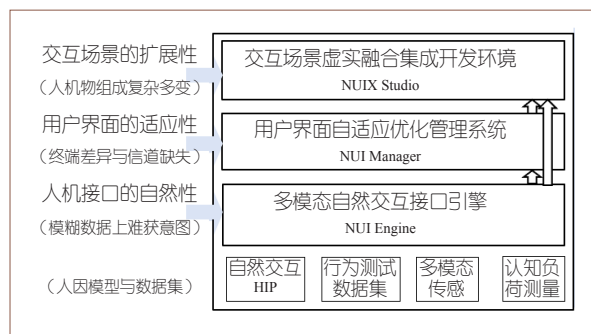


图8 自然人机交互关键技术体系NUIX (灵犀)

多用户连接进入场景进行交互体验、测试和快速迭代,虚实融合方式大大降低了场景交互的开发难度,并可在完整物理实现交互场景前进行超前验证,同时,应用众包开发技术复用更多的工程资源,显著提升场景开发的扩展性。

NUIX Studio 在教学和科研上都发挥着重要作用, 已经多年作为我教授的人机交互课程的实验平台, 特别是近两年, 由于疫情的原因, 有的课堂部分甚至全部学生都不在校园内, NUIX Studio 虚实融合的开发环境(从纯虚拟到纯物理实现)能够很好地支持学生在线合作完成实验。在科研上, NUIX Studio 上产生了一批设备互联和协同交互关键技术以及教室、研讨室等交互任务复杂的人机混合工作空间系统, 由于拓展了多用户共享交互通道, 显著增强了态势认知、提高了决策效率, 在关键应用中建设了多用户协同指挥系统, 显著提升了筹划决策能力。

我长期在人机交互领域做研究实践，基于上述 4 个层面的系统性研究成果并逐渐形成如图 8 所示

的自然人机交互关键技术体系 NUIX (灵犀), 并在教研实践中不断更新。

研究方法的形成和发展，对一个最初并不熟悉、且需要发现研究问题的领域尤为重要。在我国人机交互相对其他领域，研究社区不大，基础较弱，2005 年之前国内教研机构在顶会顶刊极少有论文发表，我带领的研究团队，2006 年陆续在 ACM CHI、UIST、Ubicomp 和 IHCS、ACM TOCHI 等发表论文。在 2021 年的 ACM CHI 上，清华大学和我本人都荣登机构和个人论文数榜首。2016 年至今，在计算机学科 CSRankings 的论文成果排名上，我也列全球第一，在技术成果上为我国高速发展的智能终端产业提供了关键技术支撑。

人机交互让机器更好地适应人，适应人的本性，适应人的操控能力、感知能力和认知能力。进一步地，我们也正在开展人机混合智能的研究，把人的作用引入到智能系统的计算回路中，以此把人对模糊、不确定问题分析与响应的高级认知机制与机器智能系统紧密耦合，使得两者相互适应，协同工作，形成双向的信息和知识的交流，形成增强的智能决策能力。



史元春

CCF 会士, 人机交互专委会主任, 2021CCF 杰出成就奖 (夏培肃) 获得者。清华大学计算机系“长江学者”特聘教授、人工智能研究院智能人机交互中心主任、普适计算教育部重点实验室主任。主要研究方向为人机交互、普适计算等。shiyv@tsinghua.edu.cn

(本文责任编辑:许嘉)