

普适计算^{*†}

吴朝晖 潘纲
浙江大学计算机学院, 杭州, 310027
{wzh, gpan}@zju.edu.cn

1. 引言

普适计算 (Pervasive Computing 或 Ubiquitous Computing) 是 1991 年由美国 Xerox PARC 实验室的 Mark Weiser 首次提出 [1] 的一种超越桌面计算的全新计算模式。他认为:“最伟大的计算技术是那些消失的技术,它们将自己融入到日常生活用品中,以至于它们从人们的视线中消失”。普适计算具有两个关键特性:一是随时随地访问信息的能力;二是不可见性,通过在物理环境中提供多个传感器、嵌入式设备、移动设备和其他任何有计算能力的设备,从而在用户不察觉的情况下进行了计算、通讯、并提供各种服务,最大限度的减少用户的介入。

普适计算这一新型的计算模式建立在分布式计算、通信网络、移动计算、嵌入式系统、传感器以及等技术的飞速发展和日益成熟的基础上,它体现了信息空间与物理空间的融合的趋势,反映了人们对信息服务模式的更高需求——希望能随时、随地、自由地享用计算能力和信息服务,使人类生活的物理环境与计算机提供的信息环境之间的关系发生革命性改变。普适计算关键在于以人为本,而不以计算机为中心,因此,普适环境中各组成元素的功能角色将不同于桌面计算模式。此时,设备是进入应用-数据空间的门户,而不再是用户必须管理的客户软件的存储库;应用是用户执行任务的途径,而不再是为挖掘设备能力而编写的软件;计算环境是一个信息强化了物理空间,而不再是为储存和运行软件而存在的一个虚拟环境。

当前计算机理论和技术难以满足普适计算模式的要求,需要发展与其相适应的计算机科学理论和技术。鉴于普适计算的重要意义,以及它带来的巨大挑战,目前已成为国际上一个蓬勃发展的研究热点。世界上主要发达国家的学术界、政府、以及工业界都非常重视,投入了大量的人力、物力、财力,各自从不同的方面对普适计算展开研究。

政府方面,美国的 DARPA、NSF、NIST 都为普适计算研究设立了庞大的研究计划,如美国国防部的研究机构 DARPA 同时资助了 5 个相关的科研大项目,分别为:MIT 的 Oxygen 计划、CMU 的 Aura 计划、OGI 和 GIT 的 InfoSphere 计划、Berkeley 的 Endeavor 计划、以及华盛顿大学的 Portolano,而美国国家标准与技术研究院 NIST 则联合各大型企业研究机构专门针对普适计算制定了详细的研究计划,并由其下属的 ITL 实验室专门负责协调、制定标准、测试等工作。欧盟也大力资助了一系列的相关研究计划,如 Disappearing Computer 计划、TEA 计划 [22] 等,并于最近启动了将耗资 1000 多万欧元的 Disappearing Computer 二期计划 [41]。英国设立了资助金额达 1000 万欧元、为期六年多的 Equator 研究计划 [40]。

工业界方面,20 世纪末 IT 业突显的疲软使 IT 巨头们认识到桌面计算已经进入了发展的平稳期,需寻求新的计算模式与进行技术革新,因此,IBM、微软、Sun、HP、AT&T、Sony、Nokia 等公司都纷纷投入巨资开展普适计算方面的研究。IBM 的 Watson 研究院专门开展了相应研究,如 DreamSpace 计划等;微软研究院也很早即开展了 Easy Living 计划,对

^{*} 本工作得到国家杰出青年基金 (60525202)、国家自然科学基金 (60533040) 和教育部“新世纪”优秀人才计划 (NCET-04-0545) 的部分资助。

[†] 本文发表在:《中国计算机科学技术发展报告 2005》,中国计算机学会文集 CCFP-0002, pp.175-187,清华大学出版社,北京,2006 年 8 月

普适计算的关键问题和应用进行研究。

普适计算是信息化社会发展进程中的又一次革新,普适计算技术必将大大改变我们未来的生活方式、工作方式。它对于我国 IT 界既是一次挑战、也是一种机遇。

2. 科学问题

普适计算环境的终极目标是实现物理空间与信息空间的完全融合,普适计算区别于传统计算技术的本质特性在于强调计算模式的“以人为本”。从这种意义上说,人类活动是在普适计算空间中实现信息空间与物理空间融合的纽带,而 Mark Weiser 所描述的计算技术消失的实质上是将人类智能转化为计算智能的结果。因此,智能是普适计算的关键,在普适计算时代,人类活动行为完全浸润于智能化的自然物理空间。支持普适计算的智能化自然物理空间有别于其他智能环境的一个关键特征在于,普适计算空间中服务于人类活动的智能实体与相应的用户相行相伴,如影相随。同时,普适计算环境区别于传统计算环境的一个重要特点是对移动的支持。普适计算研究的科学问题主要有:

2.1. 普适计算的理论建模

普适计算作为一种全新的计算模式,它横跨多个研究领域如移动计算、嵌入式系统、自然人机交互、软件结构等,具有前所未有的复杂性与多样性。**亟待一种统一完备的建模体系,准确、客观地表达普适计算所特有的普适服务“无所不在”的时空特性、“自然透明”的人机交互模式、以及普适计算“以人为本”的根本理念。**从而为普适计算系统的分析、设计、实施、部署、评估等提供多方面的理论指导,为可扩展性、可维护性、自适应性、易用性、及标准化等提供模型层面的支持。

2.2. 自然人机交互

普适计算以人为中心的特点迫切需要和谐、自然的人机交互方式,即,**能利用人的日常技能进行交互、具有意图感知能力,与传统的人机交互方式相比,它更强调交互方式的自然性、人机关系的和谐性、交互途径的隐含性、及感知通道的多样性。**普适计算环境中,交互场所将从计算机面前扩展到人们生活的整个三维物理空间,交互方式应适合于人们的习惯并且尽可能不分散用户对工作本身的注意力。和谐自然的人机交互是实现普适计算环境,使其脱离桌面计算模式的关键所在。

2.3. 无缝的应用迁移

计算实体与用户在物理空间中的高度移动性是普适计算环境的一个重要特征。为了实现能“随时随地”为用户提供透明的数字化服务,必须解决无缝的应用迁移这个核心问题,即随着用户的移动,伴随发生的任务计算必须一方面保证持续执行,另一方面任务计算可灵活地无干扰地移动。“移动计算”的研究以解决物理链路层面上可靠的移动互联为目标,而**无缝的应用迁移则需在“移动计算”的基础上,着重从软件体系的角度解决由于计算实体与用户的高度移动带来的服务级的计算移动与软件流动。**

2.4. 上下文感知

普适计算环境最终必须有自适应、自配置、自进化的能力,所提供的服务能和谐地辅助

人的工作，最大程度地减少人对行为方式及环境的关注，而把注意力集中在人所要完成的工作本身。为此，系统首先必须知道整个物理环境、计算环境、用户状态等方面的静态和动态信息，即上下文(context)。只有具有上下文感知能力，才有可能根据具体情况采取最合适的动作自主、自动地为人们提供透明式的服务。为此，需重点研究上下文感知技术，包括上下文的采集、建模、推理及融合等。上下文感知技术是实现服务自发性和无缝移动性的关键。

3. 最新发展趋势

2005 年度普适计算在方方面面都取得很大的发展，出现更多的普适计算原型系统，智能空间得到更加广泛的应用，支持普适计算的新型设备与硬件系统层出不穷，位置感知技术更加成熟，等等。综合分析本年度整个普适计算领域的进展情况，主要发展趋势总结如下：

✧ 智能空间的应用逐渐呈多尺度化、复杂化

智能空间可视为普适计算的实验床。目前智能空间研究正由物理上相对封闭的环境（如智能会议室、智能房间），逐步面向相对开放的环境（如智能购物中心、智能出租车召唤），由室内走向室外，由小规模走向大规模。而其信息交互也由单个孤立智能空间的内部交互，逐渐发展到涉及多个智能空间之间的信息交互。

✧ 大规模的个人数字化“记忆”开始兴起

近几年，如何记录与分析一个人日常生活中的每一件事，使得人们能重现往日的记忆，正成为一个研究热点，相关的研究项目急剧增多。虽然记录日常生活的所有活动需要极大的存储空间，但因目前存储能力每年约翻一倍，故这种应用的主要瓶颈不在于存储，而在于如何合理、准确地标识和过滤原始的视频、音频数据，使可以方便地浏览与检索，使系统可以作进一步的智能分析。

✧ 射频标签已成为上下文感知中最重要的设备

射频标签（RFID）是一种利用射频信号自动识别目标对象并获取相关信息的技术。它的优势在于无需复杂的计算机视觉与模式识别算法，亦无需用户作繁琐的输入动作即可识别，其设备也非常简单。可穿戴计算的顶尖会议 ISWC05 有 6 篇论文的工作都用到射频标签[2]，2006 年 IEEE Pervasive Computing 杂志第一期给出了一个关于 RFID 技术的专题，对 RFID 最新发展及其在普适计算中应用进展做了综合报道。射频标签已成为上下文感知应用中的一种通用设备。

✧ 智能手机作为普适计算的一个重要终端型载体，发挥越来越关键的作用

与其他终端设备相比，智能手机具有普及率高、用户基础好、便携性佳、计算能力相对较强、随时随地的无线通信能力等优势。全球手机数量已超过 10 亿，在亚洲的几个主要国家，手机使用率已超过固定电话。基于智能手机的高端服务可以实现许多人们日常急需的信息服务，3G 以后的智能手机已是首选的信息服务终端、普适系统的控制设备[3]。

✧ 越来越重视人机交互中的自然、和谐问题

从近两年的三个普适计算一流会议上论文情况看，学术界已越来越重视普适环境中的自然人机交互问题以及用户研究，而关于设备、硬件方面的文章逐渐减少。这说明普适计算正逐步深入，向更加核心的问题展开工作。

4. 国内外最近动态

4.1. 最新国际动态

国际上,学术界和产业界对普适计算一直保持着一贯的重视态度。当前学术界普适计算研究领域高水平的国际学术会议由原先的一个逐步发展到三个,分别为:

- ✧ **UbiComp** (International Conference on Ubiquitous Computing): 始于 1999 年
- ✧ **Pervasive** (International Conference on Pervasive Computing): 始于 2002 年
- ✧ **PerCom** (IEEE International Conference on Pervasive Computing and Communications): 始于 2003 年

其中 UbiComp[4]举办年数最长,每年的参加人数已超过五百,作为一个新兴的研究方向有如此多的参会人员,这也从一个角度反应了普适计算在国际学术界所受到的重视程度。此三个一流的国际会议一直保持着非常低的论文录用率,每年仅录用论文二、三十篇,从而保持了会议的高质量性,近几年的投稿与录用情况可见表 1。

表 1 三大普适计算会议近几年论文投稿及录用一览表

会议名称	投稿论文数	录用论文数	录用率
Pervasive 2006	178	24	13.5%
Pervasive 2005	133	20	15%
Pervasive 2004	212	26	12.3%
PerCom 2006	280	38	13.6%
PerCom 2005	233	39	16.7%
UbiComp 2005	227	22	9.7%
UbiComp 2004	145	26	17.9%
UbiComp 2003	180	27	15%

2005 年度重要纪事:

- ✧ 三大普适计算国际学术会议 Pervasive, PerCom, UbiComp 分别在德国慕尼黑、美国夏威夷、日本东京召开
- ✧ IEEE 普适计算杂志 2005 年共四期每期刊出一个专题,分别为: The Smart Phone, Energy Harvesting and Conservation, Pervasive Computing in Sports Technologies, Rapid Prototyping

4.2. 最新国内动态

近两年,普适计算已受到国内政府、学术界与工业界越来越多的重视。在学术组织与学术交流方面已步上正轨,其标志是中国计算机学会普适计算专业委员会的成立[38]和首届国内普适计算学术会议的召开。国家自然科学基金委和国家 863 计划资助了若干个相关的研究课题,尤其是国家自然科学基金委 2005 年资助了两个普适计算重点项目。国内许多院校与科研机构都成立了普适计算相关的课题组,纷纷从不同的角度对普适计算展开理论及应用方面的研究,如中科院软件所、计算所、清华大学、浙江大学、北京大学、上海交通大学、南京大学、北京工业大学等。

2005 年度重要纪事：

- ✧ 2004 年 12 月中国计算机学会理事会审议批准成立普适计算专业委员会,2005 年 6 月选举产生普适计算专委会首届主任及委员。
- ✧ 2005 年 6 月中国计算机学会 YOCSEF 举行“普适计算与智能空间”专题学术报告会。
- ✧ 《中国计算机学会通讯》2005 年第三期刊出“普适计算”专题系列文章。
- ✧ 2005 年 10 月 26 日至 28 日,首届中国普适计算学术会议在云南昆明召开。
- ✧ 2005 年国家自然科学基金委同时资助了两个以普适计算为主题的重点项目“普适计算的关键理论与技术研究”和“普适计算研究 - 手语无障碍信息服务的普适计算”,分别由上海交通大学、浙江大学、香港大学联合获得,以及北京工业大学、中山大学、中科院计算所联合获得。
- ✧ 国家自然科学基金委在 2004 年与 2005 年两年中还资助了四个面上项目,分别为:
 - 浙江大学的“普适计算中的网格接入:轻量级 Portal、动态协同及语义支持”(2004)
 - 北京大学的“普适计算中基于上下文感知的个性化智能人机交互研究”(2004)
 - 西安交通大学的“基于普适环境的上下文感知共享模型研究”(2005)
 - 浙江大学的“面向普适环境的协同编辑中若干基础问题的研究”(2005)

5. 关键技术

5.1. 理论模型

普适计算的出现对一些原有的计算机理论体系带来了新的考验与挑战。为了能有效地对普适计算系统进行讨论、分析、设计、评估,非常有必要提出新的普适计算模型。当前在这方面的研究还处于非常初步的阶段。

1) 层次模型

美国国家标准与技术研究院 NIST 提出了一个层次化的普适计算概念模型[5]。该模型参考计算机网络中的 OSI 模型,将普适计算由下而上分为五层,分别为:环境层、物理层、资源层、抽象层、及意图层。环境层和物理层表示了整个物理环境、用户生理系统及普适计算系统中的物理设备;资源层作为应用软件的基础,包括可获得的计算资源、系统可利用的用户技能等;抽象层则类似于 OSI 模型中的应用层,包括普适计算系统的应用软件 and 用户内在模型,它需要维护两者的一致性;意图层表示了应用及设备的目的或用户的目标。在该模型中,从设备角度而言,越上层抽象程度越高;从用户角度看,越上层时变特性越强。

类似地,德国学者还提出了一个普适计算的三层模型[6],由下而上分为基件层(gadget)、集成层(Integration)和普适世界层(UC World)。澳大利亚分布系统技术中心的研究人员则将普适计算归纳为四个元素:设备、用户、软件组件及用户界面[7],其中软件组件指那些能动态组成完整应用的编程部件,而用户界面则是建立在多个软件组件和设备上负责与用户交互的概念实体。

2) 智能影子模型

普适计算环境是一个开放的动态网络结构的环境,嵌入了计算、信息设备和多模态的传感器等多种异质实体,同时各种普适计算服务的形式各异、内容繁多。用户实体与计算实体之间、计算实体与计算实体之间的相互交互需随用户移动或环境变化而在不同时空动态演变。而层次概念模型仅对普适计算进行大尺度的简单建模,无法作为一个统一工具表达普适

计算环境中的高度复杂性与动态性。因此，需要一种通用模型来描述与表达对象属性，刻画和捕获对象活动，为实体互操作和用户服务提供基础。

为此，我们提出了基于物理场理论的智能影子模型[8]，它从物理空间和信息空间融合这个背景出发，根据以人为中心的不同实体之间相互作用的基本原理，为普适计算建立了一个统一模型，较准确地表达了普适计算服务“无所不在”的时空特性、普适计算用户界面的“透明”交互模式及普适计算“以人为本”的基本理念。

“场”是物理学中最经典的模型之一，如电场、磁场、引力场、光量子场等，并已被其他许多学科所借鉴，如心理场。使用物理场理论的合理性表现为：场作用的无形性对应透明性，场作用的瞬间性对应实时性，场存在的普遍性对应无处不在，场的场源为中心对应普适计算的以人为本。智能影子模型中每个人都可作为一个独立场源建立相应的体验场，描述场源对某种体验的需求度。这些体验场的数量与强度都将随着人的状态、行为等变化而自治地动态改变。体验场将对作用域内的对象发生作用，其作用结果表现为人与环境状态的改变或给人提供普适计算服务。

5.2. 自然人机交互

随着普适计算中信息空间与物理空间的融合，人与计算环境的交互将从计算机面前扩展至人们生活的整个三维物理空间，人之活动场所，时时处处皆有交互。在普适计算环境中，人机交互的目标是利用人们的日常技能与习惯进行交互，同时尽可能不分散用户对工作本身的注意力。终极目标是使得人与计算环境的交互变得和人与人之间的交互一样自然、一样方便。

人类信息传递的主要渠道为文字、语言、图像。对应地，通过纸笔交互模式、语音、以及视觉进行人与计算设备之间和谐交互正成为最有潜力的自然人机交互方式，这也是当前国际上的研究热点。笔式交互可帮助人们进行快速、自然的信息交流与沟通，而在日常生活中，更多的是听觉信息与视觉信息，它们同时可使人们获得更加强烈的存在感和真实感。

1) 笔式交互

笔式交互模拟了人们日常生活中的纸笔交互环境，可以帮助人们在保持自然工作方式的同时，轻松地进行各种信息的交流、记录与处理。笔式交互其固有的非精确性和强大的信息表达能力，非常有利于快速、自然的思想表达和交流。

笔式交互较早见于 1992 年 Xerox PARC 研制的一个原型 Liveboard[9]，而后国际上掀起了研究热潮，许多家院校、科研机构及企业纷纷对此展开深入而广泛的研究，国外如 MIT、CMU、Berkeley、日本企业 Sony、Wacom 等，国内如中科院软件所、汉王公司等。2004 年秋季著名人工智能组织 AAAI 举行了笔式交互的专题研讨会[10]。笔式交互商业化的一个里程碑是微软牵头的 Tablet PC 产品的推出，微软为此专门研发了数字墨水、数字笔、笔控操作系统，从而实现了抛弃鼠标和键盘、完全用笔进行所有电脑操作，“用笔来思考和行动”，使人机交互回归到自然的纸笔时代。

2) 基于语音的交互

听觉通道是人与计算设备进行交互的最重要的信息通道之一。在日常生活中人类的相互沟通大约 75%是通过语音来完成的，利用语音进行交互的好处在于其认知负荷较低，不需要一直占据用户的注意力，可与其他通道同时进行工作而不冲突，同时人对声音信号非常敏感、

处理速度快。基于语音的交互试图通过语音识别、语音理解、语音合成等技术，实现人与计算实体在听觉通道的和谐交互。

人机语音交互的关键在于如何实现计算设备的听觉输入和听觉输出。大致步骤如下，人说话的语音数字化后经过计算设备进行语音识别与语言理解[11]，从而让计算设备知道语音的含义，实现计算设备的听觉输入。然后可通过语音合成，用“说”的方式让人们可以听到计算实体的“意图”与“想法”，从而实现计算设备的听觉输出。

3) 基于视觉的交互

虽然人们可不利用视觉进行相互交流沟通，如，仅通过谈话或仅用文字，但人们通常更乐于面对面地沟通，原因在于，通过相互所见，沟通可变得更加丰富、更加直接。人的意图和情感通常都是通过行为以及一些形体的微小变化而表现出来，肢体语言如表情、点头、个性化手势、身体移动方式、眼神等都是人与人沟通的重要信息，这些都是人与人交互的重要组成部分。

通过视觉通道感知人的行为与意图，是自然人机交互必然要追求的目标。基于视觉的交互，重点即在于让计算设备也能感知人们通过行为、肢体语言等视觉通道表达出来的意图、情感等高层个体化信息，其远期目标是能理解人们在日常生活中的任何活动与行为，从而使人的任一动作与行为都成为交互的一部分，达到无时无刻、无处不在的实时感知。

由此可以知，基于视觉的交互其核心在于解决计算设备的视觉通道输入问题。基于视觉的交互可通过多种计算机视觉技术实现交互的视觉通道输入[39]，当前的研究热点包括：

存在与位置判别：场景中是否有人？多少人？具体位置在哪里？

身份识别：用户到底是谁

视线跟踪：用户正在看什么？

姿势识别：在语义上理解头/手/人运动的具体含义。

行为识别：用户正在做什么事情？

表情识别：用户当前的情感状态如何？高兴还是沮丧？愤怒还是哭泣？

5.3. 无缝的应用迁移

在普适计算环境中用户和计算设备都具有频繁移动的特性，如何为普适计算环境提供支持无缝的应用迁移的软件基础设施是一个亟待解决但又非常困难的问题。其目标是用户的任务可在任何时候任何地方暂停执行，无需用户进行其他主动式操作，而后即可随时在另一地方的不同设备上得到原样地继续执行。例如，用户用家里电脑工作做到任一时刻因有事要去公司，他无需主动去执行保存数据、关闭各种程序等步骤，然后其在家里电脑上的整个桌面运行现场可在公司的电脑上获得原样不变的重现，从而使用户察觉不到因工作地点、计算设备、计算环境的变化带来各种差异，而继续原来的工作。

无缝的应用迁移问题涉及软件基础设施的核心，需要深入研究普适软件的体系结构、服务构建模式、开发环境、软件编制模式等。从软件角度看，一个应用可以看作由一组服务与一组资源组成，从而必须有服务移动与资源动态绑定；另一方面，还需要移动环境下应用的运行现场保存与重构机制。

1) 服务自主发现

为了给移动的用户提供不间断、无缝的服务，需要提供服务漫游，支持在不同的物理位置以及使用不同的访问终端来获取服务资源。其中，服务的自主发现是关键，需要一种服务发现的机制为用户动态、自主地检索、发现、绑定服务。

目前工业界已有多种适合不同软件平台的解决方案,如,蓝牙服务发现协议[12]、通用即插即用简单服务发现协议(简单的底层抽象)[13]、ETF 的服务位置协议[14]、Salutation 的基于属性的服务检索[15]、Jini 的服务发现协议[16]等。特别地,Jini 提供了服务的广播、发现机制,以及动态代码移动,但 Jini 依赖于 Java 和 JDK。Surrogate[17]试图在 Jini 上改进,支持非 Java 的设备。

2) 资源动态绑定

在用户和设备的移动过程中,为实现无缝的应用迁移,需要对资源进行动态绑定。根据位置的不同可将获取资源信息的绑定方法分为四类[18]:

- ✧ **资源直接移动**:随着实体的移动,相关资源随之移动。一般而言,数据库资源不能移动;
- ✧ **资源复制移动**:随着实体的移动,相关资源被复制。特别地,在多资源的拷贝时,须注意同步修改的冲突问题;
- ✧ **资源远程引用**:随着实体的移动,原来需要的资源被修改成远程引用。需要与拥有资源的远端机器进行网络通信;
- ✧ **重新资源绑定**:随着实体的移动,使用其它的可用资源。

以上四种资源绑定策略的选择依赖运行时的状态、接入设备的属性。对于计算资源充足的移动设备,可以使用资源直接移动和资源复制规则;对于资源受限的移动设备可以使用资源远端引用和重新资源绑定规则。

3) 运行现场的重构

为了实现软件运行现场的保存与重构,目前通常的做法是通过某种机制将应用的执行状态与用户的数据相分离,并借助虚拟机。如“计算胶囊”[19]与 Collective 系统[20]。

“计算胶囊”(compute capsule)由 Stanford 大学的研究人员提出,它是一种新型的抽象机制,可用它重新构建操作系统。计算胶囊提供了一个私有的、虚拟化的、机器无关的系统资源接口,把一个用户任务运行期间的所有状态都封装起来,从而将应用状态与用户数据相分离,为应用迁移提供支持。

Chandra 等采用一种集中式管理的瘦客户端体系结构(称为 Collective)来支持应用迁移,其核心思想是采用基于缓存的系统管理机制,将应用的状态和用户数据分离,分别保存在两个不同的缓存容器中,便于应用移动到新的计算设备上。

5.4. 上下文感知

由于普适计算系统运行在极其动态和异构的计算环境中,智能实体需要能够感知环境变化,并根据动态场景自适应地调整服务状态。上下文感知计算(context-aware computing)即指能够根据上下文的变化自动地作相适应的改变和配置,为用户提供合适的服务。

能够描述环境、用户和应用程序交互过程中相关实体的状态的信息,都可称为上下文(Context),既可以是静态信息,也可以是动态信息。上下文可分为三类:计算上下文、环境上下文和用户上下文。计算上下文包括计算能力、存储能力、带宽、错误率、连接建立时间、成本、安全要求、竞争、延迟等,这些都会导致协议的变化、应用程序行为的改变、应用程序功能的改变;环境上下文包括社会条件、物理条件等,例如温度、亮度、湿度等;用户上下文包括人的标识、位置、行为、偏好等。

在上下文感知计算中,有四个主要问题:上下文信息的获取、上下文信息的建模、上下

文信息的管理、上下文信息的推理。

1) 上下文获取

如何准确地获取各种不同类型的上下文信息是上下文感知中最核心问题之一。通常,具体的获取途径依赖于具体的上下文内容,底层的上下文一般通过传感系统直接获得,而高层上下文则根据底层上下文,并结合先验知识通过推理或融合得到。上下文获取通常是指底层上下文的获取。

传感器具有分布性、异构性、多态性等特性,这使得数据收集遇到了很大挑战,问题在于:(1)如何用一种统一的方式去获取传感器数据?(2)如何把传感器数据转换,并匹配为上下文数据。

目前国际上的上下文感知系统采用的策略是进行分层处理或封装处理,使对传感器的处理对服务层透明。著名的例子包括美国乔治亚理工学院 Context Toolkit 项目和欧洲的 TEA 项目。Context Toolkit 项目 [21]是上下文感知领域一个重大的进步,该项目用 widget 对传感器进行了封装,对上层隐藏了对真正的传感器进行处理的复杂度,使得 widget 成为了可重用、可定制的基块。TEA 项目[22]由 European Commission Esprit Program 提供资助,分别由比利时的 Starlab NV/SA,德国的 TecO,意大利的 Omega Generation,芬兰的 Nokia Research 四家单位共同承担。该项目在传感器层和上下文层之间提供了一个中间层,对传感器数据做先期处理并进行抽象,使得传感器层的变化不会对上下文层发生影响。

当前上下文获取中研究最多的是如何获取设备和用户所处的位置信息。了解设备和用户的位置,非常有助于我们提供更有针对性的服务。这种对位置信息的获取被称为位置感知。位置感知目前已成为普适计算领域中一个非常活跃的研究方向,UbiComp'05录用的22篇论文中有8篇是关于位置感知[4],占三分之一,其热门程度可见一斑。当前已有多种技术与设备解决不同条件下的位置测定与位置信息的跟踪,如使用红外线、无线射频、超声波、计算机视觉技术等[23]。

2) 上下文建模

上下文的多样性导致它们都有各自的表达方式与不同程度的内在联系,如何对这些繁复庞大异构的信息建立起统一的抽象逻辑模型,从而使得这些信息都能很容易被表达、推理和共享,这是上下文感知中另一个非常重要的问题。目前不同上下文感知系统所采用的建模技术各有不同[24],主要有以下几种:

表一 上下文建模方法列表

模型	特点	例子
名值对模型	最简单的一种数据结构,易于管理,但对于复杂数据结构无能为力	Schilit 等[25]、Capeus[26]
标记模型	在属性和内容上加上标记标签的层次数据结构,经常用其他标签递归定义	CC/PP 扩展[27]
图结构模型	非常易于求得 ER 模型,从而有利于在关系数据库中结构化	ORM 扩展[28]
面向对象模型	利用面向对象技术的封装性和重用性,以解决普适计算中上下文的多变性。要访问上下文信息必须通过指定的接口	TEA 项目 [22]

基于逻辑的模型	上下文被定义为事实、表达式和规则。易于推理	Formalizing Context[29], Multicontext 系统[30]
基于本体的模型	本体论是定义概念和相互关系的工具,优点在于容易达到语义上的共享	ASC 模型[31], CoBrA 系统[32]

针对易于表达、推理和共享这三个评估指标,表中这几类建模技术各有所长,也各有所短。对上下文的建模方法,目前学术界尚未有统一的共识和标准。

3) 上下文存储与管理

鉴于上下文的广泛性,随着上下文应用的不断扩大,普适计算系统中的上下文数据必将变得很大,对这些数据需进行合理的存储,并加以一定方式的管理,从而能以最快、最佳的方式供应用程序与用户访问。

在这种移动、分布式的计算环境中,上下文数据存储的主要问题有:数据放置、数据缓存和缓存数据置换。1) 数据放置:数据存储一般放在相联的节点上,这些节点之间形成层次;2) 数据缓存:把数据缓存在移动客户端,当客户收到一个查询时首先从本地缓存中搜索。对于维护上下文感知系统下缓存的一致性,目前已经出现了位置相关的缓存有效性维护研究,通用的解决办法是在缓存数据时同时缓存其有效范围;3) 缓存数据置换:由于移动设备的缓存空间有限,数据置换技术非常必要。传统置换技术一般置换访问概率最小的数据。在上下文感知系统下还有另外两个因素:移动客户端当前位置和数据的有效范围之间的距离和数据的有效范围。一般来说,数据距离越大或者有效距离越小,被置换概率越大。

4) 上下文推理

尽管每一时刻上下文感知系统都能从传感器网络中获得大量信息,但是这些信息只是原始未加工的信息,是对环境某一属性在一个时间点上的描述,同时某些传感数据也存在一定不确定性与不可靠性[33],这使得我们难以直接获得需要的语义信息。为此,必须通过上下文推理,对这些原始信息进行信息抽取,获得语义知识,依照一定的规则进行推理,得到高层上下文。

当前上下文感知系统采用的推理技术,吸取了人工智能领域的技术,把上下文作为一种知识来推理。目前主要有以下两大类推理技术:

(1) 基于规则的逻辑推理:这是最传统也是最简单和最常用的一种推理,有很多系统采用了这种技术,主要代表为 GAIA。随着近年来本体理论的升温,语义 web 技术逐渐被应用到各个领域,包括上下文推理。基于本体的上下文推理一般都采用编写推理规则的方式。通常用 RDF 描述元数据的数据模型,用 XML、OWL 语句进行建模,以基于描述逻辑或者 FOL 编写的规则进行推理。

(2) 基于机器学习的推理:包括贝叶思网络、神经网络和基于马尔可夫模型的学习。贝叶思网络是近年来非常广泛使用的一种技术,尤其在 2000 - 2005 期间被广泛用于上下文领域,非常适合处理信息的概率分布,而且已被证明在很多应用领域内有用。神经网络技术也是上下文感知系统中使用非常多的技术,包括上下文识别和预测。

5.5. 隐私保护

在普适计算环境下,个人信息和环境信息高度结合,一部分智能设备正从人们的周围环境转移到人的身上,人们将很难把自己从信息空间中分离出来。人们所说的、所做的、甚至

个性化的感觉与情感都将会被数字化并存储起来。为了提供无干扰式、智能化、透视的个性化服务，普适计算系统必须要收集大量与人们活动相关的上下文，如偏好、当前活动状态、未来计划安排等。这使得普适计算中的隐私和安全问题的研究显得极为突出。

如何在设计普适计算系统时为隐私保护提供系统框架级别上的支持，是一个非常重要的问题。瑞士联邦工学院（ETHZ）的研究人员在深入分析普适计算的特性之后，提出了普适计算应用设计的六条指导意见[34]，并构建了一个隐私感知原型系统 pawS[35]：

- ✧ **声明原则**：关于数据收集行为的声明。不仅要声明数据的收集方式，如收集来的数据放在哪儿（即用户到什么地方去取）等等，而且要使用户能够方便的得到这些信息（如这些策略的声明等）。可借鉴 W3C 制定的 Web 隐私信息使用规范 P3P[36]；
- ✧ **可选择原则**：要提供一种选择机制，让用户可以选择哪些是愿意提供给别的用户的，哪些数据他们同意被收集或被公开；
- ✧ **匿名或假名机制**：让用户不用担心因发布信息而被跟踪或者被识别。目前的一些通信方法很容易暴露用户的身份，需对通信协议进行一些调整，如不使用固定的硬件地址。该机制的缺点在于缺乏个性化信息；
- ✧ **位置关系原则**：要根据用户与感知设备或被感知对象的位置远近来决定对相应数据的访问权，从而加强访问约束。是当用声明原则解决问题遇到困难时的一种方法；
- ✧ **增加安全性**：由于引入了一些新的约束条件（如有限的能源、计算能力等），使现有的一些安全技术不能直接用于普适计算环境中，有必要对其进行重新审视与研究；
- ✧ **追索机制**：应具有追索用户所拥有的数据被使用的情况，相当于数据使用的日志，以满足法律的要求。

随着人们对生活质量要求的提高，隐私保护问题解决的好坏必将直接影响到普适计算的发展。为更好地解决普适计算环境中的隐私保护问题，最近欧盟甚至还特别制定了欧洲隐式计算机（Disappearing Computer）的隐私设计指导方针[37]。

6. 面临的挑战

随着研究的不断深入，普适计算正面临着更深层次的技术需求。围绕人活动行为的“随时随地”与“以人为本”是普适计算理念中最关键的两点。随着用户活动的时空范围的扩大，普适计算环境将表现出前所未有的动态性、异构性以及多样性、设备的种类繁多、彼此功能和能力各异、网络通信协议多样、资源状态动态变化等等都要求系统能够自动做出适当调整以继续服务。因此，用户的高度移动性将对普适计算带来多重挑战：

- （1）迫切需要一种比智能空间更为自然、有效的、具有高度流动性的普适计算模型，在该信息化生活服务环境中，应直接以人为中心，在任意时刻，所需服务都会像影子一样以最佳方式出现在用户身边；
- （2）上下文的建模及其感知目前还处于比较浅的层次，而对时序、空间、多尺度上下文的建模并进行深入的加工处理，提供对用户高度流动性的支持，是当前极为迫切的问题。如何对上下文深层次多尺度的建模，使用新的分析组织使用手段，以便更好的减少普适计算中对用户的干扰，是当前极其重要的研究课题；
- （3）人机交互理论和技术需要突破。目前的计算技术不是以人为中心的，其交互方式不是自然的。普适计算从根本上改变了人使用计算机的方式，要使计算机服务于人，提供新型的、行为感知理论和技术具有重要意义，在这一方面，目前的研究成果还

不能满足需要。

- (4) 为满足普适计算环境中对用户无缝的应用迁移、无线、Ad-Hoc 方式、上下文驱动和非连续性服务 QoS 保障等方面的要求，迫切需要构建一种自适应的、高度动态、强可扩展性的、支持无缝移动性的、更加有效的软件基础设施。

普适计算已成为继桌面计算后的又一次计算模式的革命。欧美各国的政府、学术界、工业界都已高瞻远瞩地投入了巨大的人力、财力进行相关研究。但是与国外普适计算的科研水平相比，我国尚有很大差距，原因是多方面的。第一，普适计算本身是建立在移动计算、无线通讯、嵌入式系统、人机交互、数据管理等众多领域基础上的一种全新的计算模式，基础要求较高，尤其需要多学科交叉与合作；第二，普适计算也是实践性很强的学科领域，它需要一定的产业基础，需要与工业界的交流与合作；第三，普适计算所涉及的一些关键技术需要较为长期深入地研究，更需要一些有远见的有力度的科研投入。变革也是一种机遇，我国的计算机界和产业界有必要、也完全可能在此领域中做出高水平的研究成果，从而在普适计算时代到来时掌握主动。

参考文献

- [1] Weiser, Mark. The computer for the 21st Century. Scientific American 265(3): pp. 66-75. September 1991.
- [2] ISWC'05, <http://www-static.cc.gatech.edu/gvu/ccg/iswc05/>
- [3] 吴朝晖，“智能手机”，中国计算机学会通讯，第三期，2005。
- [4] UbiComp, <http://www.ubicomp.org>
- [5] A. Dima, L. Ciarletta, A conceptual model for pervasive computing, International Workshops on Parallel Processing, pp.9-15, Aug. 2000.
- [6] M. Muhlhauser, Ubiquitous computing and its influence on MSE, International Symposium on Multimedia Software Engineering, pp.48-55, Dec. 2000.
- [7] K.Henricksen, J.Indulska , A.Rakotonirainy, Infrastructure for Pervasive Computing: Challenges, Proc. of the Informatik 2001: Workshop on Pervasive Computing. pp. 214-222, Vienna, Sep. 2001
- [8] Zhaohui Wu, et al, SmartShadow: A Field-based Conceptual Model for Pervasive Computing, 2006. In preparation for submission.
- [9] S. Elrod, R. Bruce, R. Gold, D. Goldberg, F. Halasz, W. Janssen, D. Lee, K. McCall, E. Pedersen, K. Pier, J. Tang, and B. Welch. Liveboard: a large interactive display supporting group meetings, presentations and remote collaboration. In Proc. of ACM CHI '92, pages 599-607, 1992.
- [10] <http://www.aaai.org/Symposia/Fall/fss04.php>
- [11] Neal Alewine, Harvey Ruback, S. Deligne, "Pervasive Speech Recognition," IEEE Pervasive Computing, 3(4):78-81, 2004.
- [12] <http://www.bluetooth.com>
- [13] <http://www.upnp.org>
- [14] SLP. RFC2608
- [15] <http://www.salutation.org>
- [16] <http://www.jini.org>
- [17] <http://surrogate.jini.org>
- [18] P. Bellavista, A. Corradi, R. Montanari, C. Stefanelli, "Dynamic Binding in Mobile Applications: a Middleware Approach. " IEEE Internet Computing, Vol. 7, No. 2, pages 34-42, Mar.-Apr. 2003.
- [19] B. K. Schmidt. Supporting Ubiquitous Computing with Stateless Consoles and Computation Caches Ph.D.

Thesis, Computer Science Department, Stanford University, August 2000.

- [20] R. Chandra, N. Zeldovich, C. Sapuntzakis, M. S. Lam. The Collective: A Cache-Based System Management Architecture. In Proceedings of the Second Symposium on Networked Systems Design and Implementation, pp. 259-272, May 2005.
- [21] Anind K. Dey, Daniel Salber and Gregory D. Abowd, A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications, Human-Computer Interaction (HCI) Journal, 16(2-4):97-166, 2001.
- [22] TEA: <http://www.teco.edu/tea>
- [23] Jeffrey Hightower, Gaetano Borriello, Location Systems for Ubiquitous Computing, 34(8): 57 – 66, IEEE Computer, August 2001.
- [24] Thomas Strang, Claudia Linnhoff-Popien: A Context Modeling Survey. Workshop on Advanced Context Modelling, Reasoning and Management, joint with UbiComp 2004, Nottingham/England, September 2004
- [25] Schilit, B. N., Adams, N. L., And Want, R. Context-aware computing applications. In IEEE Workshop on Mobile Computing Systems and Applications , Santa Cruz, CA, US, 1994.
- [26] Samulowitz, M., Michahelles, F., And Linnhoff-Popien, C. Capeus: An architecture for context-aware selection and execution of services. In New developments in distributed applications and interoperable systems, Kluwer Academic Publishers, pp. 23–39, Krakow, Poland, September 2001
- [27] Indulska, J., Robinsona, R., Rakotonirainy, A., And Henriksen, K. Experiences in using cc/pp in context-aware systems. In LNCS 2574: Proceedings of the 4th International Conference on Mobile Data Management (MDM2003), Lecture Notes in Computer Science (LNCS), Springer, pp. 247–261. Melbourne, Australia, January 2003
- [28] Ted McFadden, Karen Henriksen, Jadwiga Indulska, and Peter Mascaro. Applying a disciplined approach to the development of a context-aware communication application. In 3rd IEEE International Conference on Pervasive Computing and Communications (PerCom), pages 300-306. IEEE Computer Society, March 2005.
- [29] J. McCarthy, Notes on formalizing contexts. In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, pp. 555–560, San Mateo, California, 1993
- [30] Ghidini, C., And Giunchiglia, F. Local models semantics, or contextual reasoning = Locality + compatibility. Artificial Intelligence 127(2): 221–259, April 2001
- [31] T. Strang, Service Interoperability in Ubiquitous Computing Environments. PhD thesis, Ludwig-Maximilians University Munich, Oct. 2003.
- [32] Harry Chen, "An Intelligent Broker Architecture for Pervasive Context-Aware Systems", PhD Thesis, University of Maryland, Baltimore County, December 2004.
- [33] A. Ranganathan, J. Al-Muhtadi, Roy H. Campbell, Reasoning about Uncertain Contexts in Pervasive Computing Environments, IEEE Pervasive Computing, 3(2):62-, 2004
- [34] Marc Langheinrich, Privacy by Design – Principles of Privacy-Aware Ubiquitous Systems, UbiComp'01, LNCS-2201, 2001.
- [35] Marc Langheinrich, A Privacy Awareness System for Ubiquitous Computing Environments, UbiComp'02, LNCS-2498, 2002.
- [36] P3P 1.1 specifications, <http://www.w3.org/P3P/>
- [37] Saadi Lahlou, Francois Jegou, European Disappearing Computer Privacy Design Guidelines, version 1.1, October 2004. <http://www.rufae.net/privacy.html>
- [38] CCF: <http://www.ccf.org.cn/>
- [39] M. Turk, "RTV4HCI: A Historical Overview," in B. Kisanin, V. Pavlovic, and T. Huang (eds.), Real-Time Vision for Human-Computer Interaction, Springer, 2005

- [40] Equator Project: <http://www.equator.ac.uk>
- [41] Disappearing Computer: <http://www.cordis.lu/ist/fet/dc-sy.htm>