

48-50

多通道用户界面中的目标选择技术*)

Object Selection Technology in Multimodal User Interface

方志刚¹ 马卫娟² TP311.5

(浙江大学 CAD&CG 国家重点实验室 杭州310027)

(浙江大学电子工程系 杭州310028)¹ (浙江大学外语学院 杭州310028)²

Abstract In multimodal user interface the sufficiency of input information is more important than the accuracy, and the users are allowed to input information through inaccurate modalities in human-computer interaction process. Aiming at the requirement for implementation of fuzzy object reference, the paper presents an algorithm of object selection through inaccurate input information in multimodal user interface, applying fuzzy pattern recognition theory.

Keywords Multimodal user interface, Multimodal integration, Reference technology, Object selection, Fuzzy pattern recognition

1 引言

国内外近年来兴起的多通道用户界面研究强调^[1]: (1)多个通道以并行、协作的方式工作; (2)以充分性代替精确性, 通常允许不精确的输入信息。其中, 允许不精确输入进入人机交互过程是多通道用户界面的显著特征之一, 它意味着可以消除不必要的精确性, 从而大大降低用户的认知负担。然而, 目前国内外提出的多通道整合算法^[1,2]在处理手段和算法实现方面基本上沿用传统的精确方式, 使得问题和方法配合方面存在一些不足。本文提出利用模糊数学理论解决多通道用户界面整合中目标指称这一关键问题, 可以更好地实现对不精确信息进行整合。

2 多通道用户界面及其整合算法

2.1 多通道用户界面的表示模型

通过对传统用户界面与多通道用户界面的分析和比较, 可用一个三元组 $M=(N, P, R)$ 来统一表示传统的和多通道用户界面(图1), 其中, M 表示多通道用户界面, N 表示通道(或输入设备)的数目, P 表示输入精确性, R 表示通道间关系。若 $(N>1)$ 是多通道用户界面传统区别于传统用户界面的前提特征; P 既受到当前交互设备的精度的限制, 又考虑到为降低用户认知负荷避免不必要的精确性; R 关系到人机交互的自然

性和效率, 例如人总是习惯于手势与语言并用地与他人交流信息。

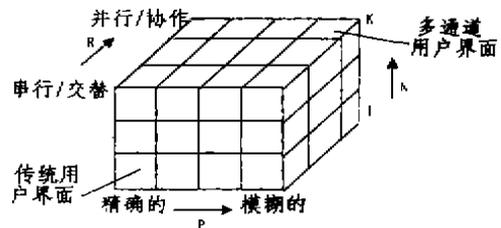


图1 多通道用户界面的三维表示模型

2.2 多通道用户界面的直接操纵特征

多通道用户界面本质上是直接操纵方式, 这是由它所追求的目标和特性决定的。因为直接操纵用户界面已被人们广泛接受, 它的成功在于其“所见即所得”特性, 充分挖掘了人的形象思维的长处而减少由于逻辑思维所带来的认知负担。直接操纵界面所基于的用户行为模式为 SSOA (Syntactic/Semantic Object-Action, 即语法/语义的目标-动作模式^[3])。Microsoft Windows 95 的设计者称为“以文档为中心的界面”^[4], 是以面向对象为思想基础的。在这种模式中, 用户只关心文档而不是动作或程序; 命令的语法结构是“名词-

*) 本文得到国家自然科学基金和浙江省自然科学基金项目的支持。方志刚 博士, 副教授, 主要研究方向为人机交互、软件工程等。

动词”，而不象命令语言系统中的“动词-名词”结构。

2.3 多通道指称技术分类

根据直接操作用户界面“名词-动词”模式可知，目标选择是多通道用户界面中基本的和核心的操作。一般而言，基本交互任务包括^[5]：(1)改变视点；(2)选择目标；(3)操纵目标。毫无疑问，“选择目标”是实现任何其它操纵任务的前提，“操纵目标”的对象都是针对“本地目标”的，其中“改变视点”主要作用是辅助目标选择（将感兴趣目标滚动到视野中）。目标选择可以使用不同的指称技术。

指称(reference)技术无论在人际交往还是在人机交互中都广为存在。基本的指称技术通常有：①描述式，如“长度为25，宽度为15的矩形”；②直指式(directic)，如用手指着屏幕上的矩形并说“这个”；③指代式，如“它”可指代曾经提到的矩形。可是，传统人机交互技术中使用的指称技术不仅单一，而且基本是基于精确方式的。本文研究主题就是实现一种有效的基于模糊方式的指称技术。

3 模糊目标选择算法

Pausch^[6]使用“属性空间”距离判断来实现不精确输入的目标选择，该方法将具有不同量纲和范围的属性特性参数归一化后，求得与用户输入的未加工的输入流的一组“描述属性”值的加权或不加权的距离，并选择其中最小距离者作为匹配目标。这种方法存在的问题是：(1)从某种意义上说它是一种“绝对”选择，因为照这种方法通常总有目标被选中，即使跟用户的意图是明显不合理的，例如在所有“大的”长方形中选择“小的”长方形。原因在于它对人的自然表达包含的日常知识未予关心；(2)通常“描述属性”（如用户说“大的”）是离散值，而“目标属性”（如目标大小）却是连续值，此时如何归一化“描述属性”呢？只有牺牲“描述属性”的模糊性，将其“量化”为确定值从而演变为精确匹配。由于模糊数学理论能够将用户日常经验和知识反映到交互过程中，因此可望借以解决多通道输入的模糊匹配问题，避免机械的数学变换。

3.1 模糊模式识别原理

目标选择可被看作针对属性参数所作的模式识别问题。存在两种情形的模糊模式识别^[7]：(1)最大隶属度原则Ⅰ。确定目标 X_1, X_2, \dots, X_n 中哪一个最接近模糊集 A ；(2)最大隶属度原则Ⅱ。确定目标 X 属于模糊集 A_1, A_2, \dots, A_n 中的哪一个。识别算法的一般步骤包括：(1)特征抽取和模式生成；(2)隶属函数建立；(3)识别判决。常用原则有：(a)最大原则；(b)阈值原则；(c)混合原则。其中(b)适用目标复选，而(a)，(c)用于目标单选。运用模糊模式识别原理的目标选择选用

最大隶属度原则Ⅰ算法，即以用户对目标的“描述属性”作为模糊子集，分别计算场景中所有目标隶属于该模糊子集的隶属度，并选择最大者。以下讨论有关算法的实现细节。

3.2 特征选择和提取

目标的选择可分为应用无关的和应用相关的两类特征。应用无关的特征主要有：空间位置、尺寸大小、名称、位序等。应用相关的特征如颜色、形状、重量、密度等。现以应用无关特征为例进行讨论，如：(1)位置模糊集：如“上，下，左，右，左上角，中央”，等等；(2)大小模糊集：如“大，小，中等大小”等等；(3)名称模糊集：名称本为精确而唯一的特征，但自然语言中常有些不确切的表达如拼写错误，不完全名，缩写以及重名等；(4)位序模糊集：传统用户界面中目标的位序是确定的，如用Tab键遍历目标，或用箭头键遍历菜单项，而自然的交互风格应支持对用户使用模糊词汇如“下一个”，“上一个”，“前面的”，“后面的”，“附近的”进行遍历；(5)颜色模糊集：包括日常生活中常用的各种颜色，通常只使用有限的几种，这对于非专业应用领域已经足够了；(6)直指模糊集：这是一类非语言的属性模糊子集，来自指点手势或视线等输入通道，反映用户在交互场景中的“感兴趣区域(AOI, Area Of Interest)”。

3.3 隶属函数的确定

模糊数学理论关于隶属度的确定有多种方法，包括统计法、主观评价法等。我们以几个应用无关特征为例子以说明：(1)位置：位置特征的隶属函数可采用以目标的中心点为自变量的分段函数，分别考虑二维和三维空间的情形。首先对交互场景或当前视窗进行分割，并找出特定位置（如“左上角”）的典型坐标，然后计算目标中心点与此位置的几何距离，并以该距离为自变量选择一种标准的隶属函数（如 $f(x) = e^{-k(x-a)}$ ）；(2)对于远远偏离该位置的（如“右下角”）目标，则令 $f(x) = 0$ ；(2)形状大小：隶属函数的确定与位置特征相似，先找出参照标准，再确定目标的实际指标与该标准值之间的函数关系。例如，对于“大的”、“小的”等特征模糊集，比较的指标可以是几何形体的面积（二维）或体积（三维）；对于“长的”、“方的”、“扁的”等特征，比较的指标可以是长宽比或宽高比等等；(3)位序特征用离散模糊集，如“附近的”，可表示为 $A_{near} = \{(-1)/0.9 + (1)/0.9 + (-2)/0.7 + (2)/0.7 + (-3)/0.4 + (3)/0.4 + (-4)/0.1 + (4)/0.1\}$ ，其中 $(-i)$ 表示前面的第 i 个， (i) 表示后面的第 i 个；(4)颜色：颜色的名称是人们约定俗成的，虽然每种颜色在色度图上占据一点位置，但对人的视觉而言实际是一个范围，颜色在该范围内的变化对人的视觉是等效的（林仲贤[1987]；杨治良[1990]），而一般人在日常生活中能分辨的颜色数目更

是有限,我们可以在 CIE 色度图上为典型的颜色选取一个中心点,然后计算目标的颜色(一般用 RGB 颜色系统表征)与该中心点的欧氏距离,并以该距离为自变量选择一种标准的隶属函数;(5)直指:手势或视线输入不能象鼠标器那样精确地指点设备精确控制到屏幕像素一级,只能反映具有一定范围的所谓“兴趣区域”,而且这个范围的界限是模糊的。假设“兴趣”所指具有一定的概率分布(如为正态分布),反映了可能存在一个兴趣中心。

3.4 “描述模式”生成

由于用户描述目标属性时常常组合各种属性,有时会上加修饰词如“最”,“很”,“比较”等,因此根据“描述属性”生成“描述模式”需借助模糊集合运算,包括:(1)求并(or), $C=A \cup B$,两种属性取其--即可;(2)求交(and), $C=A \cap B$,两种属性同时具备;(3)求补(not), $C=A^c$,不具备指定属性;(4)平方, $C=A^2$,加强属性;(5)开方, $C=A^{1/2}$,削弱属性。

3.5 指称范围的确定

目标选择可以有单选和复选两种,视不同交互任务而不同。三种典型的目标选择范围是:①特定单选,要求描述无歧义,使用“这个”、“那个”(“the”、“this”、“that”)等限定词,或使用“最”(“most”)等修饰词;②非特定单选,允许描述存在多义性,可从中选取任一满足条件的目标即可;③复选,使用“这些”、“每个”、“所有”(“all”、“those”、“these”、“every”)等限定词。

3.6 识别判决算法

目标选择的过程是根据用户的语言(或文字)描述的属性或手势指点的方位等模糊信息确定当前场景中符合这些属性及其组合的目标。根据不同的交互任务可有单选和复选两种形式,我们采用“最大隶属度原则 1”完成目标识别算法,从交互场景的可达目标中找出与用户描述最佳匹配的目标。具体步骤如下:

(1)根据用户描述的属性信息分析并选取有关特征,形成目标模式(target pattern),根据“描述属性”中的形容词确定模糊集;根据程度修饰词确定对该模糊集的基本隶属函数的变换,如集中运算(平方)、散漫运算(开方)等;根据连接词确定不同模糊集之间的运算,如并(或)、交(和)、补(非)以及笛卡尔积等;根据范围限定词确定判决类别为单选或复选。由于允许用户组合各种属性,而为不同属性建立的模糊子集属于不同的论域,因此必须求出这些属性模糊子集的笛卡尔积。对于用户没有提及的属性则予以忽略,换言之,该属性不对本次识别提供约束信息;

(2)根据目标模式的隶属度函数计算交互场景中每个目标在“描述属性”中所涉及的属性模糊子集的隶属度;

(3)根据“描述属性”中包含的逻辑关系(与、或、非)和程度修饰词(很、比较等)计算目标的组合隶属度;

(4)若为复选操作,则使用“阈值原则”判决,选择出隶属度高于阈值的所有目标;

(5)若为单选操作,则综合使用“阈值原则”和“最大原则”判决,即在要求隶属度高于阈值的前提下,选择隶属度最大的目标。在单选情形下,有可能发生选择失败,其原因可能是由于输入的描述信息不充分(上溢),也可能由于不存在满足约束属性的目标(下溢)。

在上述算法中,计算“复合隶属度”要考虑一些因素。因为对于模糊目标选择而言,无论单选或复选,其结果都附有一整合概率(置信因子),特别地,精确选择的置信因子应为 1.0。该置信因子作为后续信息整合的概率输入而参与概率推理,因此在识别算法过程中计算而得的“复合隶属度”未必可以作为最终的置信因子,而只能对筛选目标有效。为了消除歧义,用户可能附加许多限制属性,这导致“复合隶属度”趋于减小。然而对于最终的选择而言,一旦解决了所指,其置信因子则不应太小,而应使用某种归一化的置信因子。

结束语 本文是国家自然科学基金重点项目“多通道用户界面研究”的相关研究内容,在该项目中我们提出一种基于概率变换的多通道整合算法^[1],试图统一整合精确的和模糊的输入信息。可用性测试表明,通过适当地确定隶属函数可改善选择结果的合理性,可更好地符合用户的交互意图。

致谢 感谢浙江大学心理学系主任、博士生导师王坚教授对本文的帮助。

参考文献

- 1 Chen M, et al. Task-Oriented Synergistic Multimodality. In: Proc. of the First Intl. Conf. on Multimodal Interface (ICMI'96). Beijing, 1996
- 2 Nigay L, Coutaz J. A Generic Platform for Addressing the Multimodal Challenge. In: Proc. CHI'95 Human Factors in Computing System. ACM New York, Denver, 1995
- 3 Shneiderman B. Direct Manipulation: a Step Beyond Programming Languages. IEEE Computer, 1983, 16(8)
- 4 King A. Inside Windows 95. Microsoft Press, 1995
- 5 Wang H W, et al. A PC-Based Platform for VR Development-PCVRS. In: Proc. of the Fifth Intl. Conf. on CAD/CG'97, Shenzhen, 1997
- 6 Pausch R, Gossweiler R. Application-Independent Object-Selection from Inaccurate Multimodal Input. In: Blastner M M, Dannenberg R B, eds. Multimedia Interface Design. Massachusetts Addison-Wisley, 1992
- 7 黄崇福,王家鼎. 模糊信息分析与应用. 北京:北京师范大学出版社, 1992
- 8 方志刚. 多通道用户界面模型、整合方法及可用性测试: [杭州大学博士论文]. 1998