

32-36

人际通信 图

人机对话

软件工程 (8)
人机界面

人际通信中图的作用及其对人机对话的意义

Pictures in Human-Human Dialogue and Its Implications to Human-Computer Dialogue

刘正捷

(海事大学计算机系 大连116026)

朱宗元

(国家科委西南信息中心 重庆400013)

TP311.5

摘要 The difficulty for the construction of picture semantics oriented human-computer dialogue comes mainly from the complexity in human's understanding of picture semantics and the utility of the semantics in dialogues. The paper proposes a layered model for picture semantics and discusses the classification for pointing actions to picture semantics, attempting to provide a conceptual framework for analyzing certain types of human-human dialogue. The discussion on this basis about the human-computer dialogue suggests that (1) choices should be made for layers and dimensions in the picture semantics model according to the need of user's task and be limited to as shallow layers as possible, (2) the layers chosen determine the types of pointing action and (3) introduction of the reference context could facilitate the handling of pointing actions.

关键词 Human-computer dialogue, Picture, Semantics, Pointing action

目前人机界面之所以不能支持针对图的语义的对话,原因在于机器既不理解图的语义,也缺乏对图的语义的运用能力。相比之下,面对图的对话者形式的人际通信有两个基本要素,一是对话双方对图的语义的相同理解,这构成了双方通信的公共基础(common ground)^[2];二是双方在表达与理解信息时对图的语义有相同的运用方式。因此,在建立针对图的语义的人机对话时,必须使机器中图的语义模型

以人对图的语义理解为基础,使人机对话符合人在对话中对图的语义的运用方式。然而,人际对话中对图的语义理解和运用现象的复杂性正是建造此类系统的难点所在。为此,必须认真分析图的性质以及图在人际通信中的使用,特别是人对图的语义理解和语义运用,从中获得有关建立相应人机对话的有益启示。

本文首先从人际通信角度对图的概念作出限

定。另外为了支持复合重用构件的动态组装结构,语言要实现动态事件触发机制,这也要求能够运行时获取类型信息,本文给出的Java实现方案已经在IBM VisualAge for Java上原型实现,得到了初步的验证。

致谢 本文的研究工作得到了国防科工委、中科院和IBM中国公司的支持,作者在此深表感谢。

参考文献

- [1]刘西洋等,中文报表可视化复合重用构件的研究与开发,西安电子科技大学软件工程研究所研究报告,1997年7月
- [2]Java Beans 1.00-A. SUN Microsystems Corp., Dec. 1996
- [3]VisualAge for Smalltalk Programmer's Guide to Building Parts for Fun and Profit(Third Ed.), IBM Publication SC34-4496, 1997

- [4]Krasner G. E. et al., A Cookbook for Using the Model-View-Controller User Interface Paradigm in Smalltalk-80, J. of Object Oriented Programming, 1(3) 1988
- [5]Martin Fowler and Kendall Scott, UML Distilled—Applying the Standard Object Modelling Language, Addison Wesley Longman, 1997
- [6]John Barnes, Programming in Ada95, Addison-Wesley, 1996
- [7]ANSI/ISO Working Paper for Draft Proposed International Standard for Information Systems—Programming Language C++, ANSI X3J16/96-0018, ISO W921/N0836, Jan. 1996
- [8]Frank Buschmann et al., Pattern-Oriented Software Architecture—A System of Patterns, John Wiley & Sons, 1996
- [9]Java™ Core Reflection, SUN Microsystems Corp., 1996

定,接着根据人对图的语义理解给出图的一种层次语义模型,并讨论对图的语义的一种主要运用方式——指点引用的分类,为问题的解决提供一个初步的概念基础。在此基础上围绕人机对话探讨一些有关的问题,为系统建造提出策略性的建议。

1 图的概念

图是一种通过静态视觉模拟手段来表达信息的媒体。按照图所表达的内容,可以将图分为具象图(presentational picture)、抽象图(abstract-graphical picture)和象形图标(pictogram)三种类型^[1]。

具象图,所表现的是现实或虚拟世界中事物的可见属性和关系,与所表现对象之间有几何相似性。这种图范围很广,包括了从地图、景物图解到摄影照片在内的各种图。

抽象图,表现的是现实或虚拟世界中事物的不可见属性和关系。一般通过两种方法使不可见的属性和关系成为可见,其一是采用图形符号,如程序框图和部门组织结构图;其二是将不可见属性映射于颜色、大小、形状、质感、位置等可见属性来使之成为可见。

象形图标,介于具象图和抽象图之间,表达的是不等同于其自身的某种含义。图形用户界面上各种图标便是典型例子,它们在特定上下文中有明确含义,且往往可用文字来替换。

以上分类是一种大致的区分,实际上对具体的一幅图有时可能难以作出唯一归类,因为图的内容往往同时兼有不同特征,而且这除了取决于图本身之外,还与其用途有关。本文所针对的主要是具象图,当然这并不排除许多论点对其它种类图也同样适用。

2 人对图的语义理解:层次语义模型

迄今人们关于人对图的语义理解这一认知过程的认识仍然很不够。有关的认知心理学和语言学研究还难以对此向计算机科学提供系统的有实际指导意义的理论。因此,系统建造时往往采取经验性的和面向任务的处理方法。

研究指出^{[1][11]},图采用模拟和图形手段进行表达。通过提供视觉相似性来表现事物丰富的特征性(specificity)信息,但同时,由于缺乏对特定抽象层和理解范围的聚焦,会造成较大的联想空间和解释自由度。这与语言的聚焦和非特征性质形成对照,另一方面,图的表达兼有随意和非随意两种成分。前者不伴随和依赖表示与被表示间某个已存在的映射,除非同时引入一个新的映射系统,否则其解释具有较

大随意性;而后者由于依存于已有的映射系统,其解释具有确定性。这些性质使得图在表达上有如下特点:擅长表现视觉上可见的信息和空间性、结构性信息,但不擅长表现抽象概念、动作以及数量、时间、否定等概念和元信息,容易激发人的直觉、形象思维和情感,有利于人快速获得模式,能引起丰富联想。

图的这些性质决定了人在图的语义理解上轮廓与细节相分离的特点^[4]。图一般只以直陈方式提供其语义信息的轮廓,细节要靠一个基于上下文和背景知识的主动的细节推断和模型构造过程来获得。在一般情况下,当人看到一幅图,如一幅汽车图片时,会得到什么信息?首先,会看到汽车的外貌、各部分的位置、形状、大小、颜色及彼此的空间关系等画面上直接显现的事物;其次,可根据以往的视觉经历容易地想象出被遮挡住的另一侧的车轮以及发动机等画面上不直接可见,却在相应现实场景中可见的事物;此外,基于背景知识,可能还会联想到车的年代、用途以及诸如方向盘与导向轮的关系等画面和现实场景中都不可见的信息。由此可以认为图的语义由以下三个层次组成。

(1)视觉表现语义:这是图直接表现的信息,可凭视觉观察获得,主要包括图上直接表现对象的位置、外观及对象间的空间关系。由于图实际上是问题域空间的一种模型或表示,人对视觉表现语义的理解可横跨模型级和问题域级两个抽象层^[3],如地图上的一条直线可当作是模型级的一个直线图元,也可在问题域级当作是所表示的泛指的一条路,或现实世界中具体的某条路,这取决于对话兴趣点。因此,该语义可被分成视觉图元信息和视觉表示信息,而后者又可细分为基于约定的和基于应用域的信息,比如地图上某圆圈符号在基于约定的层次代表一个泛指城镇,在基于应用域的层次则代表某个具体城镇。

(2)视觉想象语义:这部分信息虽然未在图中直接表现,却存在于图所表现的现实视觉场景。人们可根据视觉经验,靠视觉推理(visual reasoning)从相应的视觉概念模型(visual mental model)中获得。视觉想象语义与视觉表现语义一起构成所表现对象的可见信息集合。

(3)关联语义:这部分信息未在图上直接表现,也不存在于相应视觉场景,却与视觉表现语义和视觉想象语义有密切联系,这包括图中对象的名称、非可见属性(理化、历史、文化、时间、象征等)及对象间非可见关系(力学、逻辑、功能等)。可以从相应概念

模型较容易地联想到这些信息。它们构成了所表现对象的非可见信息。

值得指出,虽然从信息性质上对图的语义作了如上区分,但人在理解一幅图时一般并不自觉意识到这三层语义的存在,而是将它们视为一个有机整体,在对话中运用图的语义时才从方式上有所区分。

3 人对图的语义运用:指点引用

人际通信中对图的语义的运用是复杂而微妙的。人们可以采取明显的运用方式,如用单纯的言语手段来引用,言语的特点使之较适于引用有确定名称的对象,或缺乏可见对应物的抽象对象。但当引用涉及空间位置或结构性信息时,言语表达往往复杂而冗长。因此,对话中对图的语义更常见的引用是用“这个”、“那里”等代词和手势来进行的,用手势在地图上的波斯湾上面勾画一个圈同时说“这里”,即完成了对波斯湾地区地域概念的引用,从而简化了言语表达。对图的语义还可用隐含或暗示的运用方式,比如在讨论某机械装配任务时,分解式装配图上部件的数量、种类及装配顺序和复杂程度等对专业人员一目了然的信息,为对话双方提供了某种预先存在的共识。这与通过先前对话达成的共识在对话中可起同样作用。再比如,讨论城区行车路线时,交通图上街道的分布、连通关系及标明的交通管制规定在对话中也可有类似作用。在明显和隐含两种语义运用方式中,后者似乎在对话中起着更大更深的的作用,同时也更加复杂。对图的语义运用现象的准确认识还有待进一步研究。这里只对属于明显语义运用方式的语义运用现象——指点引用进行讨论。

引用(reference)在人际对话中不可缺少,对话涉及的对象必须靠引用来引入对话。指点引用(pointing action)^[9]是一种很常见的引用方式,典型例子是说话人用手势指向一所房子,同时说“这所房子…”。指点引用由互补的两个成分构成,一是非言语成分,它给出所引用对象的空间位置,可通过身体姿态、目光凝视或面部表情,但最常见的还是通过指向手势来实现;二是相伴的言语成分,用来对所引用对象给出类型、数量等方面的限制,为指点引用的解释提供框架。指点引用发生的一个前提条件是有一个对话双方共同可见的视觉上下文(visual context),一般是当时所处的可见空间,但在许多面向任务的对话中,视觉上下文只是该空间的一个子集,这由任务性质所决定。以某幅图作为视觉上下文在面向任务的对话中是普遍存在的。这里手势指向的视觉上下文中的对象为呈示物(demonstratum)、

言语成分称为描述成分(descriptor),整个指点引用所引用的对象称为引用物(referent)。使用指点引用的目的是将听话人的视觉注意力引向说话人欲引入的对象,以实现对视觉上下文信息的引用,简化对话的言语部分,便于理解。从轮廓与细节相分离通信的观点,这样是把直接的轮廓传递移向间接的细节传递^[5],从而提高通信效率。

指点引用有着多变的运用方式,以实现视觉上下文信息多层次的引用。按引用物、呈示物、描述成分和视觉上下文间的关系,指点引用可分成标准指点引用和非标准指点引用两个基本类型,而后者又可进一步细分。

3.1 标准指点引用

这种指点引用中引用物是可见的,存在视觉上下文中,且指向手势和言语成分用来指出同一对象,即呈示物就是引用物。这里的言语和非言语成分互为补充,前者对引用物的某种属性加以描述,后者则给出引用物的空间位置。用指向手势指向一个人,同时说“这个人…”是典型的标准指点引用,引用物即是那个人。视觉上下文有时充当着另一问题域空间的模型或表示,如地图或机械图解等,这时指点引用的运用必须注意语义层次的存在,明确区分模型本身与它所表现的问题域空间,如指向地图上的一条线说“这表示河流”(模型空间),或者说“这是长江”(问题域空间)。

3.2 非标准指点引用

在非标准指点引用中,呈示物并不等同于引用物,其作用是为了引出引用物。这种指点引用有许多变形。

1) 引用物存在视觉上下文中 (a) 引用物与呈示物共存于视觉上下文;对象间存在局部与整体关系时,可指向一个对象(局部)而引用包含该对象的另一对象(整体),如指着车轮的轮胎部位说“这个车轮”时,引用的是整个车轮而不是其中的轮胎部分。引用物还可是呈示物的属性或涉及的过程,如指向一件家具说“这样式”,指向一个比赛中的运动员说“那个运动项目”。(b) 使用移位指向手势(shifted pointing):在说话人使用接触式指向手势时,有时为不遮挡引用物而指向引用物旁边或下边的位置。此时手势直接指向的对象与引用物并无内在联系,听话人必须熟悉这一手势使用规则才能准确识别引用物。

2) 引用物不存在视觉上下文中 呈示物可充当代表非可见引用物的符号,说话人常通过在视觉上

下文中给出相应可见符号将不可见的对象引入对话。这里可区分两种情况。(a)同类指点引用:呈示物与引用物属同一类型。可指向某对象而引用该对象所属类型的所有对象,如指向某棵树说“这种树生长慢”。还可指向某对象而引用同类的另一对象,如指向某人身上的衣服说“我有一件这样的衣服...”。(b)转喻指点引用:呈示物与引用物不属同一类型。通过指向代表某对象的约定符号来引用该对象,如指向某国国旗说“这个国家”,指向某人名字说“那个人”。有时呈示物是代表引用物的自然符号,这种情况下呈示物与引用物间一般有着类似于‘原因-后果’、‘财产-主人’或‘位置-对象’等内在关系,如指同一辆撞毁的汽车说“那场车祸”,指向一把被遗忘的雨伞说“这人真粗心”,指向某人刚坐过的座位说“他马上会来”。

人在对话中常用四种手势,即直指手势(deictic)、象形手势(iconic)、符号手势(symbolic)和哑剧手势(pantomimic)^[6]。对图的语义进行指点引用主要采用前两种。直指手势通过指向空间中一点或圈出空间中一个范围来指出对象,快捷而简便。象形手势是以手势勾画出对象外观某种特征来指出对象,勾画出对象外形轮廓是其常见形式,有较好的指向准确性。

4 关于人机对话

前面关于人际通信的论述对人机对话的建立是有益的启示,但那种试图让机器成为与人同样的对话伙伴的想法在技术上是不可行的,也往往没有必要^[7]。鉴于人在人际或人机对话中的对话行为有受对话伙伴影响的倾向^[7],作者主张为减少系统建造难度,应基于对人际通信的模型化和技术支持有效性分析,在保证功能需要的前提下,将人机对话引向技术上可行的方面。下面将在前边关于人际通信现象分析的基础上,探讨人机对话中图的语义模型建模及语义运用处理问题。

4.1 图的语义模型的建立

建立图的语义模型时应考虑到,人对图的语义理解固然有稳定和一致的一面,但在一定程度上会因人因事有所变化。首先,语义理解的侧重点会受用户任务影响,从事不同任务时会自然地在图上关注不同信息。如绘制机械图纸时,绘图人员关注的是点、直线等图元信息;装配机械时,装配工人在图上关心的是各部件的外形、材料及相互空间关系和装配顺序;而讨论机械操作和运行时,关注点又会移向各部件动力学和控制上的功能和关系。再比如,在一

幅地图提供的有关城市间空间关系信息中,讨论公路网规划时,人们关心的是城市间的邻近关系;而在比较各地气候差别时,会转而关心城市间纬度上的相对分布及是否靠近海洋等因素。一方面,这种由任务造成的语义理解差异,会表现为对语义层及语义层中不同方面(dimension)取舍的不同。另一方面,不同任务对浅层语义比对深层语义会有更多认同,使这种语义理解差异有着随语义层加深而扩大的倾向。

考虑人机对话时通常假定双方有协同完成同一任务的共同目标。然而,在谈论同一话题即同一任务情况下,双方对一幅图语义理解的不一致也经常存在。比如在一幅汽车图片前讨论其构造,尽管一个汽车技师和一个普通人对图上视觉表现语义的理解可能并无区别,但显然前者会从图中更快地得到更多视觉想象语义和关联语义。又如,从某城市街区图上,一个孩子和一个成年人获得的关联语义会有较大差别。这种语义理解的不一致是由于其对概念模型的依赖造成的。图的语义理解实际上是以视觉图元信息为轮廓的关于视觉表示信息、视觉想象语义和关联语义的细节推断过程,必须以相关事物的概念模型为基础的。不同的知识背景和经历使不同的人对同一事物的概念模型存在差异,这导致对同一幅图语义理解的不同。这种语义理解上的差异往往造成人际对话的失败,需要额外的澄清和修复对话来解决。应当注意到,各语义层对概念模型的依赖程度是不一样的,层次越深依赖越大。这使得不同人在语义理解上的差异呈现随语义层由浅入深而扩大的倾向,即对浅层语义的理解较易趋同,人倾向于认同可见的东西,对深层语义则容易有分歧。这似乎表明,为了向用户任务提供针对性的支持,减少因图的语义理解差异造成对话失败,降低系统建造的难度,应在图的语义建模时,根据用户任务需要对语义层次及层次中的方面作出取舍,并尽可能限制在浅层语义。同时还可考虑在界面上通过言语暗示对用户

在图上的语义关注点作必要引导。

当前机器视觉技术对图的语义理解还远达不到与人相当的水平^[7]。对图的语义建模来说,以人对语义信息的主观判断和取舍为主导的交互式建模方法似乎是目前唯一可行的途径。图的语义建模既涉及浅层图元信息,也涉及深层抽象概念,有许多难以模型化和机械处理的成分。为此,需要对图的层次语义模型进一步细化,发展将人的直觉能力和抽象概念处理能力结合于机器对图元数据的处理能力的人

机协同的建模方法和工具。人机界面上的图有两个来源,一是由图形技术产生,二是由扫描或捕获而来。前者已有图元级甚至更高的语义概念,且已从任务角度对语义信息作了过滤,若能找到将已有语义信息纳入建模过程的方法,则有助于图的语义建模。后者的表示形式是位图,语义建模需从基本的图元开始,但不涉及已有数据与建模的兼容。与此有关的另一问题是图的维数,图的语义特别是空间语义复杂度一般随维数而增加。由于经常有随任务需要对图降维使用的情况,通过省略图的维数来过滤语义信息有时是减少语义建模难度的可用方法。

4.2 图的语义引用处理

人机对话中图的语义引用的实现取决于能否找到准确高效处理引用的方法。分析表明,图的语义层次与指点引用类型间有大致对应关系。标准指点引用的引用物限于视觉表现语义。非标准指点引用中,引用物存在视觉上下文时涉及的是视觉表现语义;而引用物不存在视觉上下文时,涉及的是视觉想象语义和关联语义。这意味着根据任务需要对语义层次的取舍,将对指点引用类型产生间接限定作用,使语义引用处理得以简化。在多种指点引用并存时,应从界面技术上加以区分,向用户提供选择手段。

根据认知经济原理,对话中对引用的处理都只考虑当时直接有关的信息。为了提高语义引用处理效率和准确性,有必要引入引用上下文(reference context)概念^[9],它可定义为图的语义集合的一个动态子集,由与当前引用直接有关的语义对象组成。其形成受多种因素控制,首先是任务因素,不同任务将引用上下文限定在不同语义层和层中的不同方面;其次是语言因素,引用中的言语成分对引用上下文中语义对象类型会产生限定;引用上下文的继承性也约束其演变,保持、缩窄、拓宽和更新四种继承关系有助于引用上下文的快速界定;此外常识对此也有重要影响。如何在图的语义模型基础上建立起综合以上因素的引用上下文机制是实现语义引用处理的关键。

语义引用处理中需考虑的另一问题是人对指向手势的使用习惯。有关研究指出^[10],当对图的语义引用涉及空间位置指定或从多个对象中选择时,人才倾向使用手势,其中定位主要用象形手势,选择多用直指手势,而对其它语义引用则往往单纯用言语表达。这意味着应从这一角度对图的语义引用类型作进一步区分,给予针对性的表达支持。研究还发现,指点引用时手势在时间上一般稍超前于言语成分,

而极少有相反情况。指点引用表达的设计应提供与此匹配的言语手势结合模式。

结束语 尽管有理由相信,针对图的语义的人机对话技术上的突破将大大改进人机通信能力,但最终目标的实现仍然艰巨。这不仅要通过多学科协同深入认识人际通信中图的运用规律,而且要解决可计算问题,发展和完善图的语义建模和语义运用处理所需的理论和技术。

参考文献

- [1] Strothotte, T. & Helms, C., *Seeing Between the Pixels*, Academic Press, Berlin, 1996
- [2] Waern, Y. et al., *Communication Knowledge for Knowledge Communication*, Int. J. Man-Machine Studies, Vol. 37, 1992
- [3] Neilson, J. & Lee, J., *Conversations with Graphics: Implications for the Design of Natural Language/Graphics Interfaces*, Int. J. Human-Computer Studies, Vol. 40, 1994
- [4] 刘正捷,人际通信对人机通信的启示:轮廓与细节相分离,计算机科学,23(4)1996
- [5] Liu, Z. & Rada, R., *A Perspective of Human-Computer Communication based on Knowledge Separation*, Intelligent Tutoring Media, 6(1)1995
- [6] Schmauks, D., *Referent Identification By Pointing—Classification of Complex Phenomena*, Proc. 16th Intl. LAUD Sym. Multidisciplinary Research on reference, 1991
- [7] Johnstone, A. et al., *There Was a Long Pause: Influencing Turn-Taking Behavior in Human-Human and Human-Computer Spoken Dialogues*, Intl. J. Human-Computer Studies, Vol. 41 1994
- [8] Rim, B. & Scharatura, L., *Gesture and Speech*, In: R. S. Feldman & B. Rim (Eds.) *Fundamentals of Non-verbal Behaviour*, New York, Press Syndicate of the University of Cambridge, 1991
- [9] Schang, D. & Romary, L., *Framing the World: Towards a Localized Spatial Reasoning*, Proc. CSNLP'94, Dublin, 1994
- [10] Oviatt, S. et al., *Integration and Synchronization of Input Modes during Multimodal Human-Computer Interaction*, Proc. ACM CHI'97, 1997
- [11] Weidenmann, B., *When Good Pictures Fail: An Information Processing Approach to the Effect of Illustrations*, In: H. Mandl and J. R. Levin (Eds.) *Knowledge Acquisition from Text and Pictures*, Elsevier, Amsterdam, 1989