

多功能感知系统中的面向 Agent 技术*

姚 郑 高 文

(哈尔滨工业大学计算机系 哈尔滨 150001)

摘要 本文将人类语言分为自然语言和人体语言,阐述了人体语言的概念,讨论了人类语言的结构与包容关系,归纳了人体语言与自然语言的信息融合模型.在该模型的基础上,利用面向 Agent 技术设计了一个多功能感知系统的框架结构,提出了 Agent 关系图表示方法,并具体实现了一种人体语言感知原型系统.

关键词 面向 Agent 技术,人体语言,多功能感知,多信息融合,Agent 关系图.

近年来,有关 Agent 的研究工作正在广泛开展,给计算机智能接口的研究工作注入了活力,即试图利用 Agent 机制使人机交互能够象人与人的信息交流一样方便、自由.考察人类的交互行为可以发现,当人与人进行面对面的通讯时不但使用包括口语(语音)、书面语言(文字)等被称为自然语言的语言通道,同时还经常使用手势、表情或动作(可统称为人体语言)来达到加强语气或传递信息的效果.显然,自然语言与人体语言同时或联合传递信息,因而研究人体语言的感知模型以及与自然语言的信息融合模型对于提高计算机的自然语言理解水平和提高人机智能接口的实用性将是有意義的.面向 Agent 技术为该研究提供了支持.

1 人体语言与多信息融合

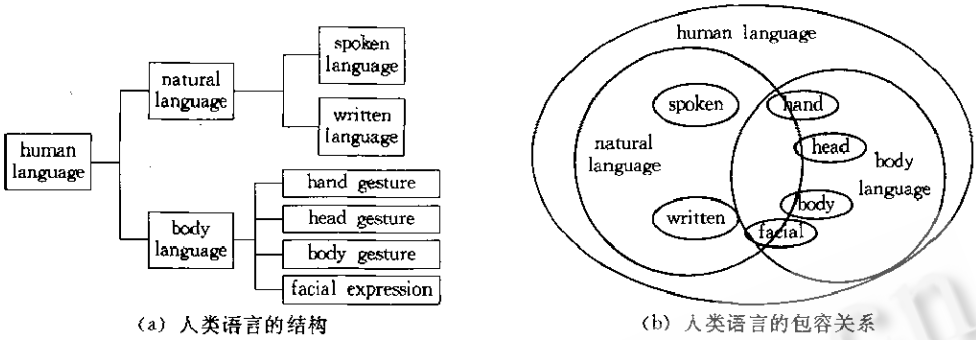
作者将人体语言定义为由包括面部表情、手势、体势、说话时所伴随的唇动,说话时所伴随的语调等功能部分组成的独立或伴随自然语言发生的一种信息传递与交换行为.^[1]如前所述,人类语言包括自然语言和人体语言,其结构与包容关系如图 1 所示.^[1]可以看到,自然语言中的语音部分与人体语言的关系比较密切,在考虑信息融合模型时这一点具有指导性意义.

显然,人类语言的感知过程是一个多通道、多信息的综合决策过程,这要求有一个很好的融合模型和规则系统.为此我们提出如图 2 所示的多信息融合模型.该模型接受 3 个通道的输入,即 1 个顺序的听觉通道,1 个并行的阅读通道和 1 个并行的(除阅读外)视觉通道.

* 本研究得到国家 863 高科技项目和国家教委跨世纪人才计划基金资助.作者姚郑,1969 年生,博士生,主要研究领域为面向 Agent 技术,软件工程和人工智能应用.高文,1956 年生,教授,博士生导师,主要研究领域为计算机视觉,多媒体技术,智能人机接口技术,虚拟现实,人工智能应用等.

本文通讯联系人:姚郑,哈尔滨 150001,哈尔滨工业大学计算机系

本文 1995-08-25 收到



(a) 人类语言的结构

(b) 人类语言的包容关系

图1 人类语言的结构与包容关系

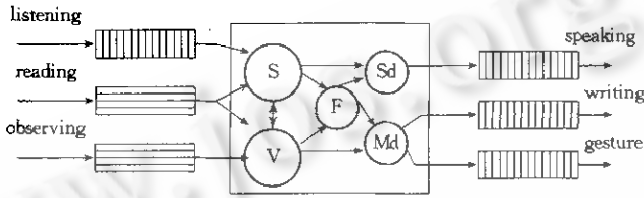


图2 信息融合的多通道模型

这 3 个通道的数据经过相应的处理模块处理后, 所得到的信息送给融合模块加工成一个信息或命令序列. 其中 S 表示声音信息处理, V 表示视觉信息处理, F 表示信息融合处理, Sd 表示声音驱动, Md 表示动作驱动.

设由自然语言流 $SF(t)$ (主要是语音部分) 和人体语言流 $VF(t)$ 融合生成的感知信息流记为 $NFL(t)$, 我们采用一种基于规则的方法来描述信息融合过程, 可用如下公式表示:

$$NFL(t) = g(SF(t), VF(t), SF(t) \otimes VF(t))$$

其中函数 $g(x, y, z)$ 是选择操作函数, $x \otimes y$ 代表两个流之间的干涉操作. 目前选定的选择操作函数如下:

$$g(f_1, f_2, f_1 \otimes f_2) = \begin{cases} f_1 & \text{if } f_2 \neq 0 \wedge (f_1 \otimes f_2) \rightarrow ET \\ f_2 & \text{if } f_2 \neq 0 \wedge (f_1 \otimes f_2) \rightarrow S - In \\ 0 & \text{if } f_2 \neq 0 \wedge (f_1 \otimes f_2) \rightarrow Fc - C \end{cases}$$

其中 f_1, f_2 分别是自然语言流和人体语言流, $ET, S - In$ 和 $Fc - C$ 表示干涉操作的规则, ET 代表“强调语气”(Emphasize Tone)规则, $S - In$ 代表“劈开插入”(Splitting and Insert)规则, $Fc - C$ 代表“矛盾删除”(Find out conflict and Cut off)规则. 在此没有考虑函数 $g(x, y, z)$ 的退化情况, 即 f_1, f_2 只存在其一, 那样就成为单纯的自然语言或人体语言的理解问题. 这 3 个规则的物理意义很容易理解, 使用 ET 规则的例子很普遍, 比如人在说话的同时伴随着手势或身体动作, 这些手势和动作只起到加强语气的作用, 语义则由自然语言表达; $S - In$ 规则要求人体语言是完整语义的组成部分, 人体语言代替了某个自然语言词汇, 比如某人说:“这个会议提供的食宿条件很……(作表示“糟糕”的手势), 讲演的情况也不过如此.”当说到食物很糟糕时, “糟糕”一词是用手势而不是用自然语言表达的; $Fc - C$ 规则也要求人体语言起到表达语义的作用, 但功能与 $S - In$ 规则不同. 比如某人说:“今天上午(摇头, 表示否定)昨天下午我告诉过大家……”, 说话人发现自己说错了一个时间时, 只简单的摇了摇头把前面的时间否定掉, 然后用自然语言说出正确时间, 这是一个典型的满足 $Fc - C$

规则的例子。

2 面向 Agent 的多功能感知系统

多功能感知系统是被设计用来研究和开发包括视觉和听觉等人类语言(自然语言和人体语言)感知的软硬件平台。我们采用面向 Agent 技术设计了该系统的框架结构。系统是由相互作用的 Agent 共同组成的,通过 Agent 之间的相互合作来实现复杂的系统功能。

2.1 Agent 及其合作类型

在系统实现时我们采用了 Y. Shoham 关于 Agent 的一个“高层次”定义,即如果一个实体的状态可被视为包含了诸如信念、承诺、能力和决定等精神状态(mental state)时,该实体就是 Agent。^[2]信念(belief)、承诺(commitment)、能力(capability)和决定(decision)等精神状态的含义与我们通常所使用的意义相似,其中决定被简单地看作对自己的承诺。这种拟人描述所带来的好处是我们拥有了一种能够用于 Agent 之间交往活动的高层通信原语,这些通信原语来自言语行为(speech act)^[3];同时这种高层描述使我们不必考虑 Agent 的内部实现问题,因而对于那些我们尚未十分了解的复杂系统来讲是非常方便的。

在多功能感知系统中,Agent 之间的合作是实现系统功能的关键。根据 Agent 之间的依赖关系,其相互合作类型可以划分为以下 4 类:

- (1) 水平型合作——每个 Agent 可以独立取得问题答案而不必依赖于其它 Agent,但如果它们之间相互合作则可以提高答案的可信度。
- (2) 树型合作——一个高级 Agent 必须依靠其它的低级 Agent 才能取得其答案。
- (3) 递归型合作——为了取得问题的答案各个 Agent 之间具有相互依赖的关系。
- (4) 混合型合作——它是前 3 种合作类型的有机结合。

2.2 多功能感知系统的框架结构

在多功能感知系统中,我们采用了混合型合作模式来体现该系统的框架结构,即该系统由许多相互合作的 Agent 组成,在总体上 Agent 按照树型合作模式相互作用,在相同层次上则采用水平型合作模式。这样,该系统的框架结构可以形象化地用图 3 表示,该图表示了系统中 Agent 的组成关系,我们称之为 Agent 关系图 ARD(agent relation diagram)。从 Agent 角度设计多功能感知系统的框架结构具有一定的优越性,ARD 表示方法较好地反映了系统内部组成成分之间的相互作用关系。

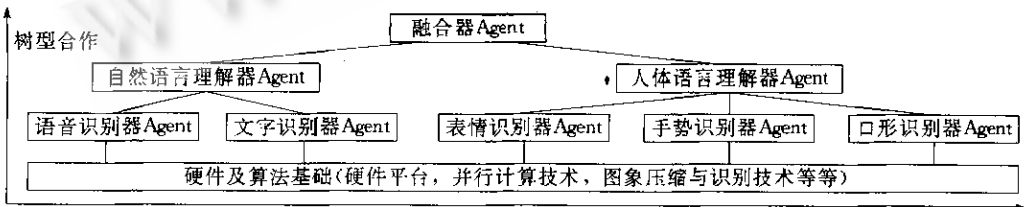


图3 多功能感知系统的框架结构

水平型合作

2.3 Agent 的构造

为了方便 Agent 的构造,我们在 Shoham 提出的 AGENT0^[2,4]的基础上设计开发了面向 Agent 的程序设计语言 AOPL, AOPL 提供了构造 Agent 的基本设施以及用于 Agent 间

通信的高层通信原语如 INFORM(通知)、REQUEST(请求)、UNREQUEST(撤销请求)和 REFRAIN(制止)等. 一个典型的 AOPL 程序框架如下:

程序开始——

能力(CAPABILITIES)部分; 初始信念(INITIAL BELIEFS)部分; 承诺规则(COMMITMENT RULES)部分;

程序结束——

能力部分定义了 Agent 的能力, 初始信念部分定义了 Agent 的初始信念, 最后的承诺规则部分是程序的主体, 它定义了 Agent 为作出某项承诺所需要的条件.

每个 Agent 总是不断地重复两类动作: 作出某个有关将来的承诺和履行先前定义的执行时间已到的承诺, 一个 Agent 程序就是承诺规则的序列, 并且伴随着事先定义的 Agent 的能力和初始信念. Agent 只能对那些它能够直接执行的动作作出承诺, 动作可以是私有的或通信的, 而且动作也可以是有条件的或无条件的. 私有动作是用 C 语言书写的子程序(它完成系统的核心功能如人体语言的识别等), 通信动作包括 INFORM、REQUEST、UNREQUEST 和 REFRAIN 等言语行为; 有条件的动作依赖于所谓精神条件(即精神状态的当前值), 而承诺的作出则同时依赖于精神条件和消息条件(当前的输入消息). AOPL 程序的解释从原理上来讲非常简单, 它由以下 2 个重复执行的步骤组成: (1) 读入当前消息, 修改 Agent 的信念和承诺; (2) 执行当前时刻的承诺, 并可能导致进一步的信念修正.

从某种意义上讲, AOPL 相当于一种高层外壳语言, 它可以用来对由 C 语言书写的源程序加以包装, 使之能够用 AOPL 的通信原语进行通信. 当然, 这种包装过程本身还存在许多技术问题, 比如如何提取 Agent 的信念和能力还没有统一的方法. 不过针对多功能感知系统的特点, 我们可以提取出相关 Agent 的精神状态. 例如融合器 Agent 要根据自然语言理解器 Agent 和人体语言理解器 Agent 所发来的信息进行工作, 这些信息之间只有时序的关系, 融合器 Agent 根据这些信息本身及其时序关系作进一步处理, 其处理依据就是前面讨论过的多信息融合规则. 在此, 融合规则被变换为谓词形式(包含时间信息)作为融合器 Agent 能力的组成部分, 其形式为(私有动作, 融合规则), 而其信念部分则是由下层 Agent(自然语言理解器 Agent 和人体语言理解器 Agent)发来的各自理解的结果(变换为包含时间信息的谓词)组成.

目前, 考虑到自然语言理解中存在着近期难以克服的困难, 我们以人体语言的感知研究为突破口, 从人体语言的识别与合成两方面进行了研究, 分别在 SUN 工作站和 SGI 工作站上用 C 语言开发出了原型系统.^[5-7]

3 结 论

本文试图以简单直观的人体语言(相对于自然语言而言)的感知研究作为研究人类智能的突破口, 以人体语言与自然语言的融合模型为基础, 从 Agent 角度设计了多功能感知系统的框架结构, 并实现了人体语言感知原型系统, 这无疑是一种很好的尝试. 进一步的工作主要集中在以下 4 个方面: (1) 软件环境与硬件平台的设计与实现; (2) 人体语言认知中的识别问题; (3) 人体语言认知中的合成问题; (4) 系统集成.

参考文献

- 1 高文. 多功能感知机的框架结构. 第二届中国计算机智能接口与智能应用学术会议, 山东威海市, 1995.
- 2 Shoham Y. Agent oriented programming. *Artificial Intelligence*, 1993, **60**:5~192.
- 3 Searle J R. *Speech acts: an essay in the philosophy of language*. Cambridge University Press, Cambridge, 1969.
- 4 Torrance M, Viola P. The AGENT manual. Technical Report STAN-CS-91-1389, Department of Computer Science, Stanford University, Stanford, CA, USA, 1991.
- 5 Gao Wen. Enhancement of human-computer interaction by hand gesture recognition. In: CHI'95 Workshop on User Interface by Hand Gesture, Denver, June 1995.
- 6 高文, 金辉. 面部表情的分析与识别. 第二届中国计算机智能接口与智能应用学术会议, 山东威海市, 1995. 243~248.
- 7 Zhang Jing, Gao Wen, Chen Xilin. A text to facial emotion and speech synthesis system. to appear on proceeding of ICYCS '95, 1995.

AGENT ORIENTED TECHNOLOGY IN MULTIPLE FUNCTIONAL PERCEPTIVE SYSTEM

Yao Zheng Gao Wen

(Department of Computer Science and Technology Harbin University Harbin 150001)

Abstract In this paper, human language is regarded as consisting of nature language and body language, the conception of body language and the construction of human language are introduced, then the fusion model of body language and nature language is put forward. Based on the model, the framework of a multiple functional perceptive system is designed using agent oriented technology and expressed as an ARD (agent relation diagram), and an implementation of prototype system for body language perception is given.

Key words Agent oriented technology, body language, multiple functional perception, multiple information fusion, Agent relation diagram.