

面向室内服务的中文语音指令深层信息解析系统^①

孔令富^② 高胜男^③ 吴培良

(燕山大学信息科学与工程学院 秦皇岛 066004)

摘 要 针对室内服务机器人的人机交互问题,对中文语音指令进行了深入研究,提出了一种基于概率/神经网络混合模型的深层信息解析系统。该系统由指令解析模块和深层信息提取模块组成,前者基于概率模型解析语音指令的有效信息,后者依据家庭环境神经网络模型,将有效信息中的服务对象或目标对象作为已知条件提取指令深层信息,旨在将指令所蕴含的深层信息显性化。构建了一般家庭条件下的实验环境进行了仿真实验,仿真数据验证了指令解析模块和深层信息提取模块的可行性;选取两类典型结构的中文语音指令,在该系统上进行深层信息解析实验,提取了准确的有效信息和深层信息。

关键词 概率/神经网络混合模型,指令解析模块,有效信息,深层信息提取模块,深层信息

0 引 言

近些年,随着人工智能和计算机技术的不断发展与进步,室内服务机器人逐渐步入人们的日常生活,代替人完成相应的家务,并协助老弱病残等行动不便的人实现生活自理。现阶段,机器人并没有实现完全的智能服务,只有接收人发出的指令(语音、手势和表情等)并且识别和理解该指令后才可以完成服务任务。因此,面向室内服务的人-机器人交互研究和应用变得越来越重要。

2008 年比尔盖茨提出了自然人机交互(natural user interface, NUI)的概念。自然人机交互是近几年人机交互领域的研究热点。依据人与人的交互形式可知语言是人们日常交流的工具,因此,基于语音的人-机器人交互是最理想且自然性高的交互方式^[1],其目的是让机器人通过识别和理解而“听懂”人类的语言,包含两个关键技术:语音识别和语义解析。语音识别技术已经相对成熟,在无外界干扰的前提条件下其识别率趋于 100%,然而语义解析领域仍处在发展阶段,且语义解析主要研究语音指令到表层语义的解析。按人与机器人的交互程度可将

语义解析分为基于人-机器人对话的语义解析和机器人自主语义解析两类。文献[2]提出了一种直接利用自然语言进行机器人运动技能训练的方法,即利用自然语言构建机器人基本行为的模糊神经网络模型,由此机器人可以通过人机对话解析简单的中文语音指令。文献[3]提出了一种通过人-机器人对话学习环境知识的方法,其引入联合概率模型实现语义解析以及语义要素到物理实体之间的映射,为机器人理解语音指令提供知识基础。机器人自主语义解析无需人机对话,它直接将语音指令解析为机器人可理解的操作语言。文献[4]利用组块分析方法提高了机器人对路径自然语言的理解,将其直接解析为机器人的运行路径。2006 年, Kate 和 Mooney 提出了基于核函数的鲁棒性语义解析(kernel-based robust interpretation for semantic parsing, KRISP)系统^[5],其将串核函数应用到支持向量机分类器中,实现英文自然语言到标准语义的解析。2011 年,文献[6]将 KRISP 系统进行了改进,提出了新的语义解析器并应用于机器人导航指令的解释。总体来说,目前语义解析主要专注于语音指令与表层语义一一对应的解析,缺乏人机交互的自然性和对指令深层语义信息相关内容的研究,为了室内服务机器

① 国家自然科学基金(61305113)和河北省自然科学基金(F2012203199)资助项目。
② 男,1957 年生,博士,教授;研究方向:智能信息处理,智能控制,计算机视觉,机器人学等;E-mail: lfkong@ysu.edu.cn
③ 通讯作者,E-mail: gaoshengnanlily@163.com
(收稿日期:2014-05-21)

人能够准确地理解指令,深层语义解析是未来该领域的研究重点。针对语义解析存在的问题,本文提出了一种基于概率/神经网络混合模型的中文指令深层信息解析系统。它由基于概率模型的指令解析模块和基于神经网络模型的深层信息提取模块两部分组成,分别利用知识库和全息地图实现了有效信息解析和深层信息提取。仿真实验和系统实验验证了该方法的正确性与可行性。

1 系统结构

本文提出的中文语音指令深层信息解析系统的深层信息解析包括两个层面:有效信息解析和深层信息提取。有效信息包括动作对象、服务对象和目标对象,深层信息包括服务对象深层信息和目标对象深层信息。该系统的基本思想是:首先,人发出的中文语音指令基于隐马尔科夫模型进行识别,其结果显示为文本信息;其次,该文本信息作为指令解析模块的输入,基于概率模型解析出文本中各个词与有效信息的映射关系,即各个词的角色标注;最后,将服务对象和目标对象作为深层信息提取模块的输入,并基于神经网络模型提取相应的深层信息。系统结构如图 1 所示。

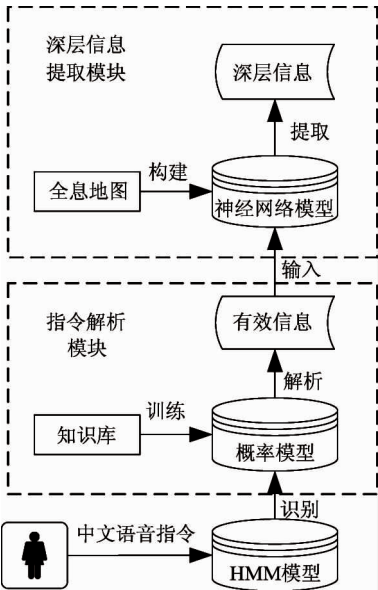


图 1 深层信息解析系统结构

从人类思维角度出发,在一般家庭环境中室内服务机器人接收的语音指令由动作、给予服务的人和操作目标组成,其中给予服务的人可以为空。本文主要研究 A 类和 B 类两类中文指令结构。A 类指

令:动词 + 人称代词“我”/人名 + 物品名称; B 类指令:动词 + 物品名称。

例 1: A 类指令:“给我一瓶矿泉水”。

例 2: A 类指令:“给小红一个苹果”。

例 3: B 类指令:“打开电脑”。

以 A 类中文指令“给我一瓶矿泉水”为例,在该系统中拟实现的解析结果如图 2 所示。图 2(a)中识别结果作为指令解析模块的输入,图 2(b)解析出动作对象[给]_{action}、服务对象[我]_{person}和目标对象[矿泉水]_{target},并将服务对象和目标对象作为深层信息提取模块的输入,图 2(c)提取出服务对象深层信息[高胜男]_{我, DI}和目标对象深层信息[冰箱、厨房]_{矿泉水, DI}。



图 2 “给我一瓶矿泉水”深层信息解析过程

2 指令解析模块

本文利用文献[7]的隐马尔科夫模型对语音指令进行识别,其结果作为该模块的输入,该模块目标是实现基于知识库的中文语音指令文本到有效信息映射: $e \xrightarrow{parse} I\{a, p, t\}$, e 表示中文语音指令识别结果(以文本形式输出结果); I 表示有效信息,包含动作对象 a 、服务对象 p 和目标对象 t 这 3 个特征值。

2.1 知识库构建

在一般家庭环境条件下获取知识库(knowledge base, KB),取 A 类指令 50 个语音样本和 B 类指令 20 个语音样本(依据人对人的说话习惯,获取人对机器人服务请求的语音指令样本,其要求简短易懂且包含有效信息),并手动标注每一条语音指令所对应的有效信息^[8],其数学表示如下:

$$\begin{aligned} KB &= \{(e_i, I_i) \mid i = 1, 2, \dots, n; n = 70\} \\ e_i &= \{w_{i1}, w_{i2}, \dots, w_{ij} \mid j = 1, \dots, m\} \\ I_i &= \{a_i, p_i, t_i\} \\ \{l_i \mid l_i \in I_i\} &= \{[w_{ix}, \dots, w_{iy}]_{annotation} \mid \\ &annotation \in \{action, person, target\}\} \end{aligned} \tag{1}$$

式中 e_i 表示某一条中文语音指令 i ,它由多个字

$\{w_{ij}\}$ 组成,其中 w_{ij} 表示指令 i 中第 j 个字; I_i 表示 e_i 包含的有效信息,即对指令 i 中第 x 到第 y 个字序列的角色标注 $[w_{ix}, \dots, w_{iy}]_{\text{annotation}}$, 其中角色标注 annotation 分为 action 、 person 和 target 三类,分别表示动作、人和目标。

例如,指令“给我一瓶矿泉水”的知识以 $(\text{给我一瓶矿泉水}, \{[\text{给}]_{\text{action}}, [\text{我}]_{\text{person}}, [\text{矿泉水}]_{\text{target}}\})$ 的形式存储在知识库中。

2.2 指令有效信息解析

指令有效信息解析模型^[9,10]的数学表示如下:

$$\text{Parse}(e, I) = \text{argmax}_p(I | e, KB) \quad (2)$$

由于有效信息 I 包含三个特征值:动作对象 a 、服务对象 p 和目标对象 t ,为了便于推导,默认特征值 a 、 p 和 t 之间相互独立,则

$$\begin{aligned} p(I_i | e_i) &= p(a_i, p_i, t_i | e_i, KB) \\ &= p(a_i | e_i, KB) \times p(p_i | e_i, KB) \\ &\quad \times p(t_i | e_i, KB) \end{aligned} \quad (3)$$

式中 $p(a_i | e_i, KB)$ 、 $p(p_i | e_i, KB)$ 和 $p(t_i | e_i, KB)$ 分别表示动作对象解析模型、服务对象解析模型和目标对象解析模型。

基于贝叶斯公式,动作对象解析如下:

$$p(a_i | e_i, KB) = \frac{p(e_i | a_i, KB) \times p(a_i, KB)}{\sum_{j=1}^n p(e_j | a_i, KB) \times p(a_i, KB)} \quad (4)$$

$$p(a_i, KB) = \frac{\text{count}(a_i, KB)}{\text{count}(a, KB)} \quad (5)$$

$$p(e_i | a_i, KB) = \frac{\text{count}(e_i, a_i)}{\text{count}(e, a_i)} \quad (6)$$

式(4)中, $p(a_i, KB)$ 表示动作对象 a_i 在知识库中的概率,即在知识库中动作对象 a_i 个数与所有动作对象个数的比值,由式(5)可计算; $p(e_i | a_i, KB)$ 表示包含动作对象 a_i 的语音指令 e_i 在知识库中的概率,即在知识库中指令 e_i 个数与包含动作对象 a_i 的所有指令个数的比值,由式(6)计算。

$p(p_i | e_i, KB)$ 和 $p(t_i | e_i, KB)$ 计算方法与 $p(a_i | e_i, KB)$ 相同,此处省略。

3 深层信息提取模块

基于面向对象的家庭全息地图^[11,12]中环境数据构建家庭环境神经网络(family deep neural network, FDNN)模型,该模型包含一个显层和两个隐层,其中显层是物品层、隐层是家具层和房间层。家

庭环境神经网络模型^[13,14]数学表示如下:

$$\begin{aligned} \text{FDNN} &= (v, h^{(1)}, h^{(2)}; \theta) \\ v &= \{v_i | i = 1, 2, \dots, I\} \\ h^{(1)} &= \{h_j^{(1)} | j = 1, 2, \dots, J\} \\ h^{(2)} &= \{h_m^{(2)} | m = 1, 2, \dots, M\} \\ \theta &= (W^1, W^2) \\ S_i &= v_i \cup h_j^{(1)} \end{aligned} \quad (7)$$

式(7)中, v 是显层-物品层,由 I 个神经元 v_i 构成, I 是物品对象(即目标对象)的个数; $h^{(1)}$ 为第一隐层——家具层,由 J 个神经元 $h_j^{(1)}$ 构成, J 是家具对象的个数; $h^{(2)}$ 为第二隐层——房间层,由 M 个神经元 $h_m^{(2)}$ 构成, M 是房间对象的个数; θ 是该神经网络的参数,由 v 与 $h^{(1)}$ 之间的权重 W^1 和 $h^{(1)}$ 与 $h^{(2)}$ 之间的权重 W^2 构成。其中,全息地图中物品层的物品对象 S_i 被分为两类:物品类对象 v_i 和家具类对象 $h_j^{(1)}$ 。

基于该家庭环境神经网络模型,深层信息提取模块表示如下:

$$v \xrightarrow{\text{extract}} DI \begin{cases} p. \text{ name}, v = p \\ (h_j^{(1)}, h_m^{(2)}), v = t \end{cases} \quad (8)$$

式(8)表示当输入 v 为服务对象 p 时,其将 p 作为已知条件输入该模块,并提取服务对象深层信息——对象名称 $p. \text{ name}$;表示当输入 v 为目标对象 t 时,其将 t 作为已知条件输入该模块,并提取目标对象深层信息——家具对象 $h_j^{(1)}$ 和房间对象 $h_m^{(2)}$, 分别标注为 $[p. \text{ name}]_{p. DI}$ 和 $[h_j^{(1)}, h_m^{(2)}]_{t. DI}$ 。

3.1 目标对象深层信息提取

本文将文献[14]的深度神经网络模型进行改进,降为适合室内家庭环境的三层神经网络模型,由两个受限波尔兹曼机(restricted Boltzmann machines, RBM)构成^[15,16],结构图如图3所示,其中目标对象作为RBM1的输入,RBM1的输出(即一级隐性信息)作为RBM2的输入。

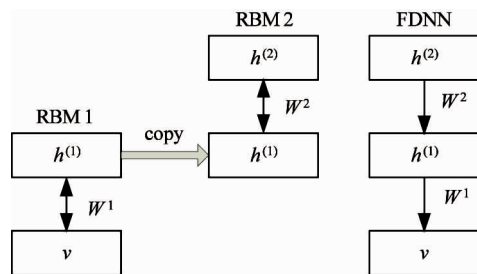


图3 FDNN模型结构图

设该模型的能量函数为 E , 如下式所示:

$$E(FDNN) = -v^T \times W^1 \times h^{(1)} - h^{(1)} \times W^2 \times h^{(2)} \tag{9}$$

式中 v 是显层, $h^{(1)}$ 、 $h^{(2)}$ 分别为第一隐层和第二隐层, $W^i (i=1,2)$ 是显层和两个隐层之间的连接权重矩阵。

基于家庭环境神经网络模型识别目标对象深层信息提取如图 4 所示。具体步骤如下:

在显层-物品层中,将目标对象作为已知条件,识别出与之匹配的显层对象 v_i ,之后依据能量函数求得目标对象的函数概率,如下式所示:

$$p(v_i;\theta) = \frac{1}{Z(\theta)} \sum_{h^{(1)}} \sum_{h^{(2)}} \exp(-E(FDNN)) \tag{10}$$

式中 $p(v_i;\theta)$ 表示在家庭环境神经网络模型中目标对象 v_i 的概率分布情况, $Z(\theta) = \sum_v \sum_{h^{(1)}} \sum_{h^{(2)}} \exp(-E(FDNN))$, 表示模型的部分函数。

在第一隐层——家具层中, v_i 作为已知条件 (即该层的输入数据), 识别一级隐性信息家具对象 $h_j^{(1)}$ 。依据条件概率分布确定家具层深层 (隐层) 信息 $h_j^{(1)}$ 的概率, 如式

$$p(h_j^{(1)} | v) = g(\sum_i W_{ij}^1 v_i) \tag{11}$$

所示,并据此一级隐性信息被确定为 $h_j^{(1)}$, 如下式所示:

$$h_j^{(1)} = \arg \max_{h_j^{(1)}} p(h_j^{(1)} | v) \tag{12}$$

在第二隐层——房间层中, $h_j^{(1)}$ 作为已知条件, 识别二级隐性信息房间对象 $h_m^{(2)}$:

$$p(h_m^{(2)} | h^{(1)}) = g(\sum_j W_{jm}^2 \times h_j^{(1)}) \tag{13}$$

与第一隐层识别方法相同, 房间层深层信息 $h_m^{(2)}$ 的概率如下式所示:

$$p(h_m^{(2)} = 1 | v, h^{(1)}) = g(\sum_i W_{ij}^1 \times v_i + \sum_j W_{jm}^2 \times h_m^{(1)}) \tag{14}$$

二级隐性信息被确定为 $h_m^{(2)}$, 如式 (15) 所示:

$$h_m^{(2)} = \arg \max_{h_m^{(2)}} p(h_m^{(2)} = 1 | v, h^{(1)}) \tag{15}$$

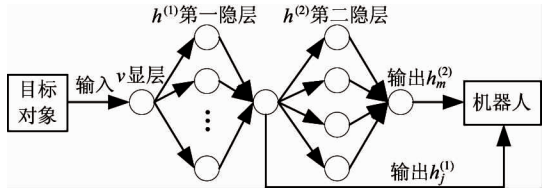


图 4 目标对象深层信息提取

3.2 服务对象深层信息提取

室内服务机器人解析语音指令的最终目标是为 人提供服务,即机器人必须知道它所给予服务的人 是哪一位家庭成员,因此,目标对象和服务对象的深 层信息提取同等重要。基于本文的两类指令结构, 服务对象深层信息提取的算法过程如表 1 所示。

表 1 服务对象深层信息提取算法

if	服务对象 p 为空,
then	$p.name = Null$, 即无特定服务对象
else	if p 为人称代词,
then	$p.name = \hat{\theta} = \operatorname{argmax}_{\theta} P(X \lambda_{\theta})$
if	p 为人名,
then	$p.name = \text{“人名”}$

该算法表示当指令中无服务对象时,则服务对 象深层信息显示为无特定服务对象;当指令中的服 务对象为人称代词,此时服务对象深层信息通过说 话人识别^[17]来判断其所对应的服务对象名称,即由 该说话人发出语音指令 X 的似然度得分 $P(X | \lambda_{\theta})$ 确定;当服务对象为人名时,其深层信息即为该人 名,在全息地图中搜索^[18,19]。

4 实验结果与分析

4.1 实验环境

在一般家庭环境条件下获取语音库,包括 3 个 说话人(2 男 1 女),分别对两类指令结构:(1) A 类 指令:动词+人称代词“我”/人名+物品名称;(2) B 类指令:动词+物品名称,进行语音样本采样(数据 采样率为 16kHz,每帧为 20ms,帧间交迭为 10ms,特 征矢量由 12 阶 Mel 频率倒谱 MFCC 及其一阶差分 AMFCC 和归一化能量及其一阶差分归一化能量共 计 26 维组成),其中包括 A 类指令 50 个语音样本、 B 类指令 20 个语音样本,每人发音 3 遍,共 630 个 中文指令语音样本。之后构建知识库,对每条语音 样本手动标注所对应的有效信息:动作对象、服务对 象和目标对象,并将指令文本和有效信息以数组的 形式存储在知识库中。

4.2 仿真实验

4.2.1 训练混合模型

由于本文语音库规模较小,为了扩大知识库的 样本数量,在训练混合模型实验时将 630 个语音样 本作为训练集,并且在实验时进行多次实验(10

次),每次实验样本数增加 1 倍,在第 6 次实验训练集为 3780 语音样本(即训练网络模型的训练集)之后,其深层信息解析准确率趋于平稳,因此,能够保证该概率/神经网络混合模型的收敛性,如图 5 所示。本文用时间复杂度 $T(n) = O(f(n))$ 计算该混合模型的复杂度,进行 10 次实验,在第 5 次实验之前深层信息解析时间随着样本数的增加而有显著提高,然而之后上升缓慢,如图 6 所示。分析收敛性及计算复杂度,本文选取 6 倍的语音样本作为混合模型的训练集。

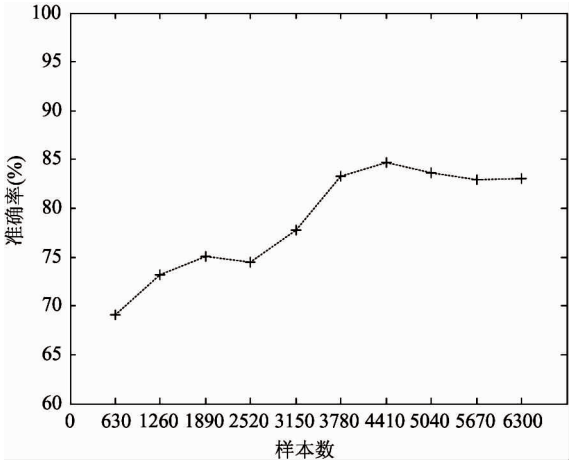


图 5 混合模型训练的收敛性

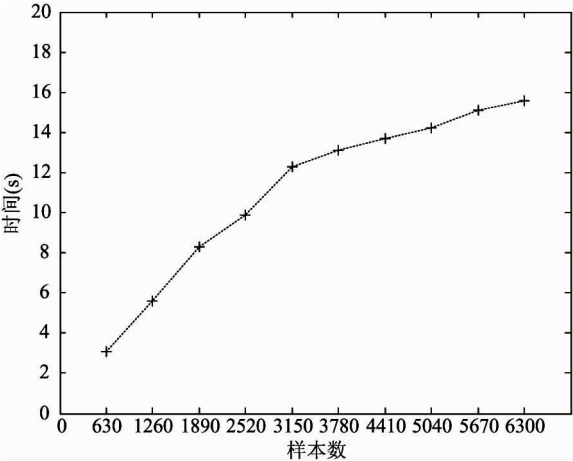


图 6 混合模型训练的复杂度

4.2.2 有效信息解析

本文随机抽取 10 个 A 类语音指令样本和 10 个 B 类语音指令样本作为测试集。在测试集上,分别利用本文概率模型、规则模型^[20]和基于核函数的鲁棒性语义解析(KRISP)模型^[5]进行 10 次有效信息解析仿真实验,其仿真结果如图 7 所示。分析该仿真数据可知,概率模型和 KRISP 模型的有效信息解

析结果相对稳定,且准确率都高于规则模型,归其原因是,规则模型适合语句相对较长的英文语音指令。但是,对于语句较短的中文语音指令,本文概率模型的准确率均值为 92.32%,比规则模型和 KRISP 模型分别提高了 6.32% 和 4.43%,且本文方法解析的有效信息最稳定。

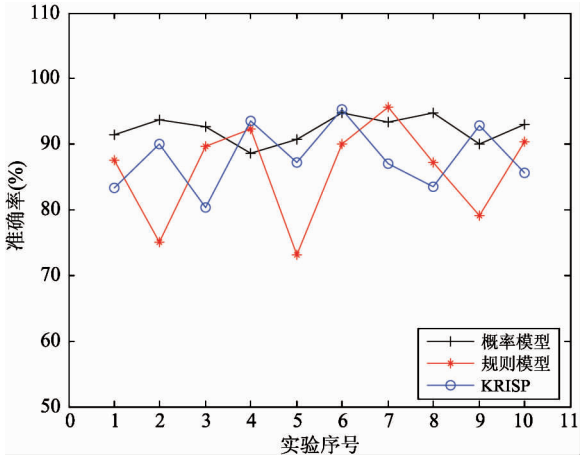


图 7 有效信息解析准确率

4.2.3 深层信息提取

随机抽取 10 个 A 类语音指令样本和 10 个 B 类语音指令样本作为测试集,基于深层信息提取模块进行 10 次目标对象深层信息提取和服务对象深层信息提取实验,A 类和 B 类语音指令深层信息提取的仿真结果如图 8 所示。分析仿真数据可知,A 类语音指令的深层信息准确率与目标对象和服务对象相关,均值为 63.63%;由于 B 类语音指令无服务对象,因此设定服务对象深层信息提取准确率为 100%,则语音指令深层信息准确率即为目标对象深

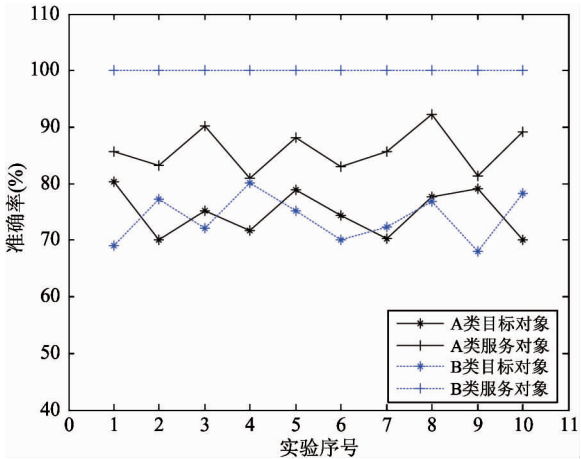


图 8 两类指令深层信息提取准确率

层信息提取准确率,其均值为73.58%。B类语音指令深层信息提取准确率比A类指令高9.95%,出现该结果的原因是深层信息提取准确率是服务对象深层信息和目标对象深层信息提取准确率之积,然而B类指令无服务对象。

4.3 深层信息挖掘系统的实例验证

以本文设计的中文语音指令深层信息解析系统为实验平台,选定某一位说话人发号两条典型的中文语音指令:A类语音指令“给我一瓶矿泉水”和B类语音指令“打开电脑”(图9)。当A类语音指令录入该系统时,启动识别(start recognize)模块识别出

文本信息“给我一瓶矿泉水”,解析(parse)模块与传统解析相似,其解析结果为有效信息动作对象“给”、服务对象“我”和目标对象“矿泉水”,然而提取(extract)模块可以提取出服务对象深层信息“高胜男”和目标对象深层信息“冰箱”和“厨房”,其解析结果如图9(a)所示。当B类语音指令录入该系统时,start recognize 模块识别出文本信息“打开电脑”,parse 模块解析出有效信息动作对象“打开”和目标对象“电脑”,extract 模块提取出目标对象深层信息“电脑桌”和“书房”,识别结果如图9(b)所示。

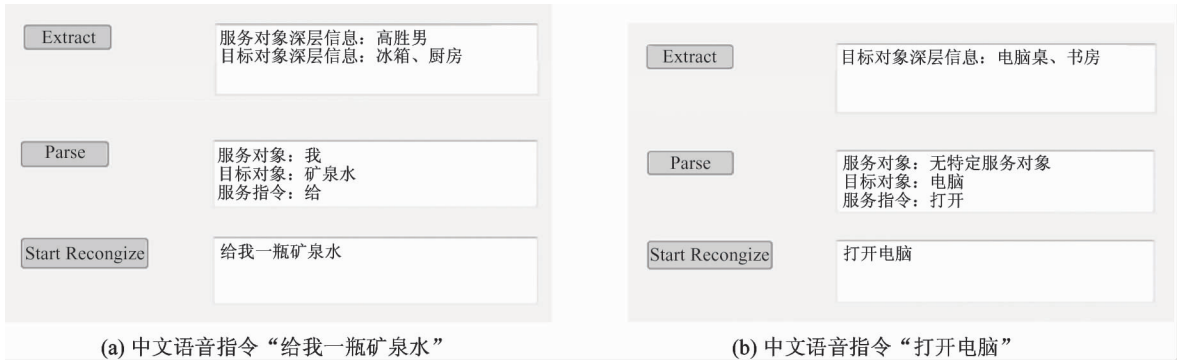


图9 典型中文语音指令深层信息解析结果

5 结 论

本文提出了一种面向室内服务机器人的中文语音指令深层信息解析系统,其创新之处是在传统解析方法基础上添加深层信息提取。该系统包括指令解析模块和深层信息提取模块,前者解析中文语音指令中的有效信息,后者提取指令隐含的深层信息,旨在为机器人准确地理解和执行语音指令提供必要的前提条件。与传统语义解析方法相比,该方法将表层语义解析提升到了深层信息语义解析,仿真实验和系统实验验证了该方法的正确性和可行性。在未来的工作中,我们将研究如何转化语音指令深层信息为机器人可执行的操作指令,以实现高智能性的自然人-机器人交互。

参考文献

[1] Daniel J, James H. Speech and Language Processing. NJ:Prentice Hall, 2000. 24-29

[2] 聂仙丽, 蒋平, 陈辉堂. 自然语言训练的机器人基本行为控制器. 机器人, 2002, 24(3):201-208

[3] Thomas K, Vittorio P, Daniele N. Learning environment

knowledge from task-based human-robot dialog. In: Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 2013. 4304-4309

[4] 李新德,张秀龙. 一种面向室内智能服务机器人导航的路径自然语言处理方法. 自动化学报, 2014, 40(2):289-305

[5] Kate J, Mooney J. Using string-kernels for learning semantic parsers. In: Proceedings of the Joint 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Sydney, Australia, 2006. 913-920

[6] David L, Raymond J. Learning to interpret natural language navigation instructions from observations. In: Proceeding of the Twenty-Fifth AAAI Conference on Artificial Intelligence, California, USA, 2011. 859-865

[7] Lawrence R. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of The IEEE, 1989, 77(2):257-286

[8] Golland D, Liang P, Klein D. A game-theoretic approach to generating spatial descriptions. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Jeju, South Korea, 2010. 410-419

- [9] Li J Z, Wang Z Y. Chinese statistical parser based on semantic dependencies. *Tsingua science and technology*, 2002,7(6): 591-595
- [10] Wong Y, Mooney J. Learning for semantic parsing with statistical machine translation. In: Proceedings of the Technology Conference of the North American Chapter of the Association of Computational Linguistics, Stroudsburg, USA, 2006. 439-446
- [11] Kong L F, Gao S N, Wu P L. Representation method and building of object-oriented household holographic map. *Computer Engineering and Design*, 2012, 34(1): 253-359
- [12] Wu P L, Kong L F, Gao S N. Holography map for home robot: an object-oriented approach. *Intelligent Service Robotics*, 2012. DOI 10.1007/s11370-012-0109-z
- [13] Geoffrey E, Dong Yu. Context-dependent pre-trained deep neural networks for large vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012,20(1):30-42
- [14] Geoffrey E, Li D. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, 2012,29(6):82-97
- [15] Geoffrey E, Simon O. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, 18(7): 1527-1554
- [16] Mohamed A, George E. Acoustic modeling using deep belief networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 2010,20(1):14-22
- [17] Huang W, Dai P Q, Li H. Text-independent speaker recognition based on classify Gaussian mixture and neural net fusion. *Pattern Recognition and Artificial Intelligence*, 2003,16(4):423-428
- [18] 吴培良. 基于智能空间中服务机器人全息建图及相关问题研究:[博士学位论文]. 秦皇岛:燕山大学信息科学与工程学院,2010. 74-85
- [19] 高胜男. 面向对象的家居全息地图建模与应用研究:[硕士学位论文]. 秦皇岛:燕山大学信息科学与工程学院, 2012. 35-36
- [20] Ge R, Mooney J. A statistical semantic parser that integrates syntax and semantics. In: Proceeding of the Ninth Conference on Computational Natural Language Learning, Stroudsburg, USA, 2006. 9-16

An indoor service system for parsing of deep information from Chinese voice commands

Kong Lingfu, Gao Shengnan, Wu Peiliang

(School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004)

Abstract

For realizing the natural human-computer interaction of indoor service robots, a study on Chinese voice commands was conducted, and a deep information parsing system based on a probability/neural network hybrid model was proposed. This system is composed of a command parsing module and a deep information extraction module, and the former parses out the effective information of voice commands based on the probability model, while the latter extracts the deep information, the service object or the target object as a condition known, based on the neural network model of family environment, for making the deep information contained in commands explicit. The experimental environment in general family environment was built for simulation experiment, and the simulation data verified the feasibility of the command parsing module and the deep information extraction module. Two kinds of Chinese voice commands with typical structure were chosen for a deep parsing experiment on the system, and the accurate effective information and the deep information were extracted.

Key words: probability / neural network hybrid model, command parsing module, effective information, deep information extracting module, deep information