

应用自动语音识别技术实现通信增值业务

Application of Automatic Speech Recognition Technology to Value-added Services

梁 烽

Liang Feng

(广州哈里斯通信有限公司语音事业部,广东广州 510663)

(Guangzhou Harris Telecom, Inc. Ltd., Guangzhou, Guangdong, 510663, China)

摘要:介绍实现商用自动语音识别的系统架构及其功能,阐述应用自动语音识别技术实现的新通信增值业务。应用自动语音识别技术可以实现语音号簿,提供个人的日程安排,事件提醒,读取邮件等,扩大信息查询内容,提供互动娱乐等通信增值业务。

关键词:通信 自动语音识别 增值业务

中图分类号:TN912.34 文献标识码:A 文章编号:1002-7378(2005)04-0309-02

Abstract: The framework and functions of the system based on commercial Automatic Speech Recognition(ASR) engine are introduced. Applying ASR technology can achieve some new value-added services such as voice phonebook, private schedule assistant, fair reminder, mail to speech, wider range information query, interactive entertainment etc.

Key words: communication, ASR, value-added services

随着固定电话网络和移动电话网络用户的普及率的提高,通信市场经营的重点从单纯的用户数量发展、用户规模经营逐步拓展至细分市场的话务量经营,电信竞争也逐步由网络资源转移到差异化的服务竞争上来。目前的通信市场已经出现了增量不增收的现象,迫切需要新的业务增长点来提高业务量和增加收入。

从目前的增值业务形式来看,最多的就是IVR系统(Interactive Voice Response,自动语音应答系统),各种信息台、点歌台、互动游戏都依托于IVR系统。现有的绝大部分IVR系统依赖于传统的人机交互方式,采用的是按键回应方式,过程冗长缓慢,繁琐而容易出错。而且按键数量也限制了输入的范围,越来越不适应信息量每日剧增的现代社会。因此,需要引入新的人机交互方式,来开发新的业务或发掘旧有业务潜力。

当前最热门的新一代人机交互方式是自动语音识别。自动语音识别技术就是让机器通过识别和理

解过程把语音信号转变为相应的文本或命令。语音识别是最自然的人机交流方式,真正可以做到“即说即做”。

目前的自动语音识别分为个人自动语音识别和商用自动语音识别两种。个人语音识别主要是对单个用户进行采样为基础,每个新用户都需要进行语音训练,不适合大规模用户群。商用自动语音识别是建立在对相同语音群体进行海量抽样分析之上,具有无须用户语音训练、排除噪声干扰、智能打断等特点,适应大规模用户群的使用,最适合引入到新的IVR业务中。本文介绍实现商用自动语音识别的系统架构及其功能,以及应用自动语音识别实现的新增值业务。

1 实现商用自动语音识别的系统架构及其功能

实现商用自动语音识别的系统结构见图1。从系统功能来看,系统可以分为4个部分:语音接入模块、语音处理模块、业务流程模块和数据处理模块。

1.1 语音接入模块

本模块负责接通用户的电话呼叫,然后将用户的声音传送到语音处理模块,并将语音处理模块返

回的语音文件播放给用户收听。

本模块功能由接入服务器群完成,接入服务器群包括接入交换机和语音板卡两部分。接入交换机一端通过E1 数字中继与公众语音网络相连,一般采用的是标准7号信令或ISUP信令;另一端通过E1数字中继与语音板卡相连。语音板卡将语音打包传送给语音识别模块,并播放返回的合成语音包。

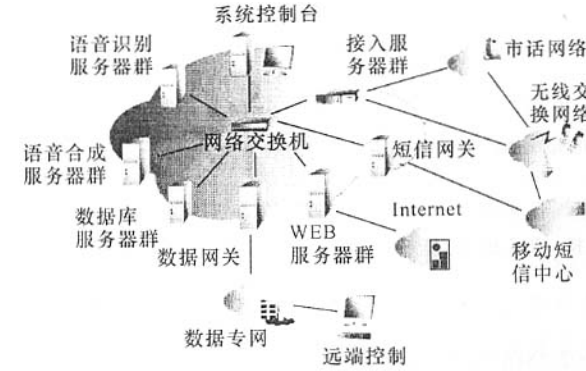


图1 实现商用语音识别的系统架构

1.2 语音处理模块

本模块识别用户的说话,提取出关键词,传给业务流程模块,并将返回的指令,合成人工语音文件,传给语音接入模块。本模块的结构见图2。

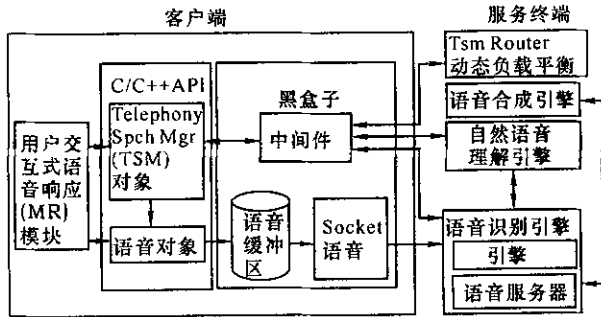


图2 语音处理模块结构

本模块功能由语音识别服务器群、语音合成服务器群完成。因为用户量大,所以本模块应该能够支持多引擎、多服务器的大型应用,应该具有可按各个机器配置不同而自动调整处理性能,支持多机分布式服务模式。

1.3 业务处理模块

业务处理是根据用户的说话,按照事先编写好的流程做出相应的反应,以实现不同类型的业务。这种交互过程的实现依赖于VoiceXML(语音扩展描述语言)^[1]。

本模块功能由系统控制台、数据库服务器群完成。系统控制台提供多种中间件服务,采用消息驱动方式,支持同步或异步请求及应答,功能包括屏蔽数据库及操作系统差异,自动平衡通信负载,系统运行

监测等。

1.4 数据处理模块

本模块实现对数据网络的外部连接和提供系统网站服务。外部的数据信息服务包括无线短信服务、信息专网连接、以及公众互连网络连接等。电信97系统和计费营帐接口也通过数据网关接入到系统中。网站服务包括系统管理网站服务和业务管理网站服务。本模块功能由数据网关、短信网关和WEB服务器群完成。

2 应用自动语音识别技术实现的新增值业务

有了自动语音识别技术的帮助,用户和系统交互的范围不再局限于小小的键盘,可以使得业务开发人员自由发挥,不仅会诞生许多新的业务,而且很多老业务一样可以焕发出新的生命力。

应用自动语音识别技术实现的新增值业务总体上可以分为:个人助理、信息查询和互动娱乐。

2.1 个人助理

自动语音识别技术在个人助理业务方面,可以充当语音号簿和生活助手。

2.1.1 语音号簿

语音号簿系统可存储大量的电话记录,能够根据人们说出的人名,自动进行电话转接或者号码查询。既免去了繁琐的按键过程,又无须记忆电话号码,使拨打电话变得更轻松的同时,带来了快捷和准确的实惠。

语音号簿可分为个人号簿和企业号簿。个人号簿为用户提供个人电话号码的查询和代拨服务。用户只需将个人的电话记录存储于号簿中,通过拨打特服号码,即可使用语音访问自己的号簿。企业号簿向企事业单位提供统一管理的号簿服务。企业将电话记录存储于企业的号簿中,统一录入、修改、增加和删除。企业的员工通过拨打特服号码,直接访问企业的号簿来查询或拨打企业号簿中的电话。外部人员拨打接入特服号,可以转接到需要找的人和部门。

2.1.2 生活助手

自动语音识别可以提供日程安排、事件提醒、读取邮件、灵感便笺等方面的业务,给个人的生活充当助手,使人们生活、工作更有效率。

2.2 信息查询

相对于传统IVR的按键输入方式,语音识别技术大大扩展了输入信息量的范围,比如:交通信息、

法就显得非常重要,它需要扩展纯定量的、面向货币的材料经济学方法,并在设计决策时认真考虑经济和人这两个重要的因素。

3.3 成本估算的准确性

影响成本估算准确性的因素主要有以下方面^[15]:(1)项目的复杂程度。虽然有一些定性的方法可用来度量工程的复杂度,但只能用在设计阶段或代码实现阶段,很难用于计划阶段。目前只能用一些主观的方法来估计一个项目的复杂程度。(2)工程规模。随着软件规模的增大,各模块之间的相关性急速地增加。(3)结构确定性程度。它是指用户需求的明确程度、功能模块划分的难易程度以及所要处理的信息的层次特性。结构的不确定性越大,估算的风险就越大。(4)可供使用的历史数据不全。(5)软件需要创造性以及人与人之间的合作,但个人或团队的行为通常很难预料。(6)软件缺乏相关的、定量的历史经验基础,小规模受控实验意义不大。(7)过去的若干年里,大量有价值的研究和数据收集已为适当准确的成本估算模型提供了一定的基础,但仍然需要进一步改进。

4 结束语

作为交叉学科,软件工程经济学既要借鉴经济学中成熟的经验、技术和方法为软件工程决策服务,也要充分认识到软件工程本身的特殊性,更要关注人与经济的协调性。尽管国内学者在该领域已经做了不少工作,但总的来说还不够,也欠系统,希望引起人们足够的重视。

(责任编辑:黎贞崇)

(上接第310页)

天气预报、街道单位、城市景点、新闻播报、百科全书等等,均可以实现。这样就能使得信息查询可以提供很丰富的内容。

2.3 互动娱乐

互动娱乐是一个很有潜力的业务类型,有望成为吸引年轻用户的业务亮点。互动娱乐包括有交友聊天、过关游戏、赛事竞猜、剧情投票等等。由于语音识别技术突破了键盘的限制,互动的内容不再局限于少数的标准答案,因此可以将模糊和类比的概念引入到其中。

3 结束语

由于电话、手机和手持移动设备(如PDA)只有很小的显示屏,有限的输入能力和有限的处理功能,

参考文献:

- [1] Webster. Webster's New Collegiate Dictionary[M]. G & C, Merriam Co, Springfield, MA, 1979. 248-250.
- [2] Gerald Weinberg. The Psychology of Computer Programming[M]. Dorset House Publishing Co, Inc, 1998. 191-192.
- [3] Boehm B. Software Engineering Economics[M]. Prentice Hall, 1981. 1-26.
- [4] 杨文龙,姚淑珍,吴芸. 软件工程[M]. 北京:电子工业出版社,1997. 253-281.
- [5] 李肯立,石归然. 提高软件成本估计精度的方法[J]. 湖南农业大学学报,1999,25(6):491-493.
- [6] 宋敬廷. 软件成本估算模型的现状与展望[J]. 水利水电自动化,1994,2:14-19.
- [7] 刘杰生. 软件成本估算方法与技术现状探讨[J]. 舰船电子工程,2000,4:13-18.
- [8] 崔敬东,左广. 软件开发成本估算方法的比较研究[J]. 技术经济,2002,11:50-52.
- [9] 晏荣杰,宋丽. 软件成本估算算法研究[J]. 华北电力大学学报,2001,28(4):61-65.
- [10] 唐敬年,宋丹峰,张怡. 计算机系统软件成本构成及估算方法[J]. 中国资产评估,2000,1:24-26.
- [11] 王平,丁浩芳,李韬. 结构型软件成本估算模型的研究与改进[J]. 计算机工程,2002,28(12):88-89.
- [12] Barry W. Boehm. Software Cost Estimation with COCOMO I [M]. Prentice Hall, 2000. 98-165.
- [13] 周杰,杜磊. COCOMOII-软件项目管理中的成本估算方法[J]. 计算机应用研究,2000,11:56-58.
- [14] 方德英,李敏强. IT项目风险管理理论体系构建[J]. 合肥工业大学学报,2003,26(增刊):907-911.
- [15] 黄成荣,胡振宇. 面向对象软件开发中的成本估算研究[J]. 广西师范学院学报,2001,18(2):30-34.

文本界面的数据处理设备很难被使用。应用自动语音识别技术实现的新增业务不存在这些限制。有了自动语音识别技术,用户可以用声音与系统交互,通过一系列的对话选择,找到所需的信息。服务器上的语音识别软件把用户的输入语音的选择转换成文本的选择,这一过程与在传统的网页上选择一个超连接类似。在电话、手机等终端设备上通过语音而不是烦琐的输入来获取信息和服务更符合人们的习惯。随着语音信号处理技术,特别是自动语音识别和语音合成技术的发展,语音将成为人机交互的一种主要形式。

参考文献:

- [1] 王建华. VoiceXML 技术语应用[J]. 电信技术,2002,(2):28-30.

(责任编辑:邓大玉)