

基于智能手机的维吾尔语语音控制系统的开发

米尔阿迪力江·麦麦提¹ 吾守尔·斯拉木^{1 2} 努尔麦麦提·尤鲁瓦斯^{1 2}
热依曼·吐尔逊^{1 2} 艾尼宛尔·托乎提²

¹(新疆大学信息科学与工程学院 新疆 乌鲁木齐 830046)

²(新疆大学新疆多语种信息技术重点实验室 新疆 乌鲁木齐 830046)

摘要 以实现维吾尔语命令词识别为目的,重点研究维吾尔语命令词识别系统在 Android 平台下的开发与实现过程,介绍系统开发难点、核心技术及系统典型的几个功能。系统主要由 Android 开发包、Eclipse 集成开发环境和 API 接口进行开发,并且通过自动选型规则来实现维汉英多种文字的正确显示及处理等问题。针对广大用户的不同说话方式,重新构建维吾尔语语音语法文件,解决各地不同方言问题。在一般实验室环境下做实验得到了 90.56% 的正确识别率和 85.00% 的成功执行率等测试结果,表明维吾尔语非特定人命令词识别研究中语法文件的结构及构建对系统有不同的影响。

关键词 Android 平台 维吾尔语 关键词识别 槽语法 命令词识别

中图分类号 TP311.1 文献标识码 A DOI: 10.3969/j.issn.1000-386x.2016.06.053

DEVELOPMENT OF UYGHUR VOICE CONTROL SYSTEM BASED ON SMART PHONE

Miradeljan Mamat¹ Wushour Islam^{1 2} Nurmatamat Yolwas^{1 2} Rayima Tursun^{1 2} Anwar Tohti²

¹(College of Information Science and Engineering, Xinjiang University, Urumqi 830046, Xinjiang, China)

²(Key Laboratory of Xinjiang Multilingual IT, Xinjiang University, Urumqi 830046, Xinjiang, China)

Abstract With the purpose of implementing Uyghur command words recognition, we elaborately studied the development and implementation process of Uyghur command words recognition system on Android platform, introduced the development difficulties, core technologies and typical functions of the system. The system was developed mainly using Android SDK, eclipse integrated development environment and API interfaces, and realised the functions of correct display and processing of multiple texts of Uyghur, Chinese and English through automatic styles selection rule. Aiming at different speaking styles of the majority of users, we rebuilt Uyghur voice and grammar files, and solved the problem of different dialects around the Region. Moreover we gained the testing results of right recognition rate of 90.56% and the successful implementation rate of 85% in the experiment made in usual Lab condition, this showed that in the research of Uyghur non-specific command words recognition, the structure and construction of grammar files had different effects on system.

Keywords Android platform Uyghur Keyword recognition Slot grammar Command words recognition

0 引言

近几年在新疆使用智能手机的用户越来越多,它将成为人们获取信息的主要设备,因此基于手机的应用软件愈来愈受到人们的关注和重视。目前 Android 技术是一个先进的、具有高人气的技术,它还是一个开放性移动设备综合平台^[1]。

我国是一个多民族的国家,新疆是个多民族地区之一^[2], Android 平台的维吾尔语手机语音控制软件一直以来都是少数民族市场上的空白。在国外,关键词识别的研究初始于 20 世纪 70 年代,那时此研究序幕由 Bridle^[3] 揭开的只称“给定词”识别,当时没有使用语法或词法信息,而是利用信号的 LPC 表示连续语音中的关键词进行了检测和定位。到 80 年代,Myers 等人^[4] 利用基于 DTW 的局部最小算法来对关键词识别和连接词识别进行研究。90 年代 MIT、CMU 和 Dragon、Toshiba 和 IBM 等公司就对 KWS 的研究得到了进一步发展,国外已经进入了高

潮,但是国内研究历史并不久。国内利用基于音节的一种汉语无限制语音流的关键词识别系统,采用了独特的统计拒识方法^[5]。科大讯飞作为国内和国际语音技术产业的领导者,国内语音技术及中文关键词识别、命令词识别技术进入了更高的一层。国内外连续语音识别及关键词识别技术取得了一定的成就,市场上也出现了一些应用产品,可是我国少数民族对关键词识别技术的研究与开发正处在初期阶段。

本文利用维吾尔语朗读式的语料训练而得到的声学模型作为本文命令词识别系统的声学模型,然后基于规则的方式,建立了槽语法文件。此文件由 15 个槽(slot)和三个语法规则(<action>)构成,再利用英国剑桥大学研发出的基于隐马尔科夫

收稿日期:2014-08-23。国家自然科学基金项目(60762006);国家工信部电子发展重大项目(159018);新疆维吾尔自治区自然科技项目(2011211A012)。米尔阿迪力江·麦麦提,硕士生,主研领域:嵌入式智能应用开发,语音处理,自然语言处理。吾守尔·斯拉木,教授。努尔麦麦提·尤鲁瓦斯,讲师。热依曼·吐尔逊,副教授。艾尼宛尔·托乎提,工程师。

2.3 槽语法

语法文件的创建在本系统开发过程中起着至关重要的核心作用。本文采用的是基于巴克斯范式 BNF(Backus-Naur Form) 的槽语法(Slot Grammar) [10]。将系统设定为只接受该语法约束下的信息查询语句,却略去无关信息,从而提高系统的性能和效率。一般槽语法中槽的个数不能太多,因为当槽的个数太多,且也有嵌套层级的情况时,将会导致产生的语法网络较复杂,直接影响语法静态扩展无法实现。

创建语法文件后利用 HTK 的 HParse [11,12] 工具得到语言模型,因此该文件的构造是整个系统的最核心技术。本系统槽语法文件是按规定维吾尔文所对应的拉丁文(如表 1 所示)书写的。槽语法由! slot,! grammar 以及 < > , [] , | 组成。除此之外,还有 < voiceAssistant > : < action1 > | < action2 > | < action7 > (三个 action 是不同的三种规则);可看出各个语法规则间是“|”即“或”的关系,还有些符号是在表 2 中给出说明。槽语法具有很强的约束性,因此系统只会接受这种语法规则下的命令词信息,且只会按照这种已规定的语法进行搜索,不能进行“回退”。主要用处就是允许语音应用程序通知语音识别器应该听到的内容,这些内容包括:可能被说出的单词、那些可能出现的单词模式和每个单词的语言等。

表 1 国际标准维文拉丁文对照表

序号	字母	拉丁文	序号	字母	拉丁文	序号	字母	拉丁文
1	ق	q	12	ھ	H	23	ز	z
2	و	w	13	س	s	24	خ	x
3	ع	e	14	د	d	25	غ	G
4	ر	r	15	ا	a	26	و	v
5	ت	t	16	ئ	E	27	ب	b
6	ي	y	17	ئ	i	28	ن	n
7	ئ	u	18	ك	K	29	م	m
8	ك	N	19	ك	k	30	خ	h
9	ئ	o	20	ل	l	31	ج	j
10	پ	p	21	ف	f	32	ئ	O
11	ز	z	22	گ	g	33	ئ	I

表 2 槽语法中各个字符的含义

grammar	语法名称
! slot	表示槽的声明。语法网络以! slot 开始,以“;”结束
! slot < >	其内的内容是槽名称
! start	定义规则起始节点槽
< >	定义规则起始节点的名称,槽为必选项
[]	槽为可选项,可内容是可选,有无对语句无影响
	表示在其左右两边任选一项,表示“或”的意思
\$	表示空格

3 软件功能实现

3.1 实现打电话功能

首先识别用户所说的语音命令后,通过 PrintContacts(c) 方法,动态地查找和获取用户手机上的联系人信息,调用 matcher (contactDisplayName) 方法来解决不符合条件的维吾尔文联系人,然后需要引用 startTalk(this, "nlp") 方法,在编写的 onButton-

RetryClick() 方法中调用命令词的识别,从而将会节省用户在一批联系人中的查找并翻阅的时间,从而更方便、快捷地完成用户打电话的需求。识别用户的命令,及转换为拉丁文,后将用户所说的语句显示在手机屏幕上,根据 mType = 1 的情况,提取槽语法中“打电话”的槽! slot < contact >;”然后以语法文件的匹配规则 [13,14] 来获取联系人。除此,当用户联系人列表修改时,可在槽语法中将会自动地更新相应的槽。系统主界面与该模块实现如图 4、图 5 所示(给“吾守尔斯拉木”打电话)。

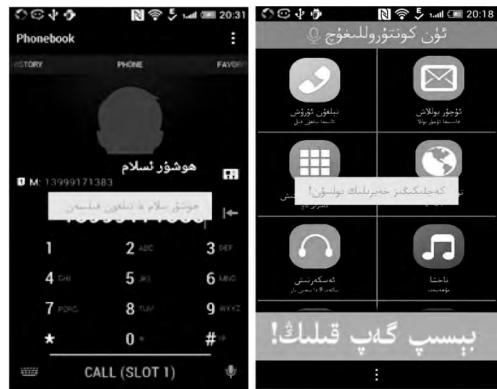


图 4 系统主界面图 图 5 打电话功能实现图

3.2 实现打开应用功能

其模块核心部分是构建应用 app.xml 文件(维文拉丁文应用名称对应映射表)在进行识别时通过识别返回的结果,参照上述 xml 文件,并且使用 Android 的 Intent() 方法打开。以往的方法是在 onResultsMsg() 方法(返回识别结果的方法)中通过 mType = 3 的整形变量来判断它的值,我们根据 app.xml 文件与 Intent 组件和应用包名,同时应用 Android 的相关类库来控制 and 实现。其所用的语法文件中对应的槽是! slot < app >;”不仅能打开系统自带的应用(相机、图库等),而且同时能打开第三方应用(QQ、微信等)。此外,不管命令词的长度和复杂度,只要是槽语法的关键槽! slot 内的词,那么都认为是命令词。因此,可知槽语法文件在整个系统中是最核心最重要的部分。其中有些应用的名称,如相机、壁纸等在维吾尔语中,有不同的称呼,本系统对这几种名称的应用程序同样会实现。而且,用户完全可以按自己想说的方式说出命令。比如对“打开相机”的测试。如:“(كامرانی ئىشلىتىمەن. كامرانی ئېچىپ بىرىڭ. كامرانی ئاچ.)”等 3 种不同的说法进行测试都能被正常启动。如图 6 所示(打开第三方应用“微信”)。



图 6 打开应用实现图

4 实验与结果分析

4.1 实验数据与环境

本文利用英国剑桥大学研发出的基于隐马尔科夫模型^[15]的HTK工具,并且在使用HTK进行特征提取、训练及得到声学模型^[16,17]的基础上,利用建立的语法文件进行了语言模型的构建^[18],再搭建本维吾尔语语音控制软件能够运行的Android开发平台。

训练集:一般环境下(如无人的教室、办公室等),录制朗读式连续语音作为训练集^[19]。发音人是18~30岁的成年人,总共356个人(189女,167男),共发声128小时的2456条语句,发音人配置高宝立式麦克风,阻抗160 Ω 、灵敏度 56 ± 3 dB,频率范围100~16 000 Hz。采样率选择16 KHz、采样位选择16 Bit。语音数据以wav文件格式存储,共有50 000多条语音文件。

测试集:录制的软件为Cool Edit 2.0,语音采样频率为16 KHz、采样位选择16 Bit、单声道格式,共有300个语音文件,即对于每一个语法规则分别录制了100个文件。语音数据以wav文件格式存储。

4.2 实验测试结果

测试环境为:系统安装至单核、RAM256、ROM256,及系统版本是Android 2.3.3的华为G606-T00和HTC智能手机,并在安静的实验室环境内进行了对于软件核心部分“意图”的人工实际测试。其中成功执行个数指的是在识别的基础上能够正确执行指令的个数,而执行失败指的是能识别但未成功执行的个数。除此,识别正确率为正确识别个数除以总数、成功执行率为成功执行个数除以总数而得。噪声等周边环境^[20]、底层语音识别率、用户声音低或者地方口音偏重、发音不够清晰正确及命令列表中不存在该词汇等不能正确识别。测试结果如表3所示。

表3 对用户“意图”执行的测试结果表

功能类别	总数	正确识别个数	成功执行个数	错误识别个数	执行失败个数	正确识别率	成功执行率
打电话	60	57	54	3	3	95.00%	90.00%
发短信	60	55	52	3	5	91.67%	86.67%
打开手机应用	60	56	52	4	4	93.33%	86.67%
上网导航	60	54	50	4	6	90.00%	83.33%
播放音乐	60	55	53	2	5	91.67%	88.33%
查询天气	60	49	45	4	11	81.67%	75.00%
总和	360	326	306	20	34	90.56%	85.00%

5 结 语

本文以维吾尔语语法的特点出发,在符合命令词的语法形式的条件下,建立了维吾尔语命令词识别的槽语法文件,通过使用HTK得到其语法网络及语言模型,并且对于一些出现的带地方口音的单词及一些新型词汇等,对最终结果有一定的影响。未来用户使用发短信功能时本系统上将增加语音输入功能,除此还需要增加语音合成功能,为了使少数民族用户使用更加方便的语音软件,让系统达到更加完美效果,对系统进行更加智能化分析、添加语音翻译器是下一步研究重点。

参 考 文 献

- [1] 韩超,梁泉. Android系统原理及开发要点详解[M]. 北京:电子工业出版社,2010:340-343.
- [2] 热依曼·吐尔逊,吾守尔·努尔麦麦提. 多文种手机混合输入/输出技术及实现[J]. 计算机工程与科学,2006,28(4):103-104,118.
- [3] Bridle J S. An Efficient Elastic-Template Method for Detecting Given Words in Running Speech[C]//Brit. Acoust. Soc. Meeting,1973.
- [4] Myers C S,Rabiner L R,Rosenberg A E. An Investigation of the Use of Dynamic Time Warping for Word Spotting and Connected Word Recognition[C]//Proc. Conf. ASSP, April. 1980:173-177.
- [5] 徐明星,郑方,吴文虎,等. 连续语音关键词识别系统的拒识方法研究[J]. 清华大学学报:自然科学版,1998,38(S1):89-91.
- [6] 陶梅,吾守尔·斯拉木,那斯尔江·吐尔逊. 基于HTK的维吾尔语连续语音声学建模[J]. 中文信息学报,2008,22(5):56-59.
- [7] Steve Young,Gunnar Evermann,Mark Gales,et al. HTKBOOK[M]. HTK Version 3.4. Cambridge University Engineering Department,

March 2009:199-211.

- [8] 那斯尔江·吐尔逊,吾守尔·斯拉木. 基于隐马尔科夫模型的维吾尔语连续语音识别系统[J]. 计算机应用,2009,29(7):2009-2012.
- [9] Wilpon J G, Lee C H, Rabiner L R. Application of Hidden Markov Models for Recognition of a Limited Set of Words in Unconstrained Speech[C]//ICASSP,1989,3(1):254-257.
- [10] Rohlicek J R, Russel W, Roukos S, et al. Continuous Hidden Markov Modeling for Speaker-Independent Word Spotting[C]//ICASSP,1989,1(1):627-630.
- [11] 李星星. 基于HMM的汉语语音关键词检测研究与实现[D]. 武汉理工大学,2009.
- [12] Rose R C, Paul D B. A Hidden Model Based Keyword Recognition System[C]//ICASSP,1990,1(1):129-132.
- [13] Christiansen R W, Rushforth C K. Detecting and Locating Key Words in Continuous Speech Using Linear Predictive Coding[J]. IEEE Trans. on ASSP,1977,25(5):361-367.
- [14] Alan L Higgins, Robert E Wohlford. Keyword Recognition Using Template Concatenation[C]//ICASSP,1985,1(3):1233-1236.
- [15] 郑方. 连续无限制语音流中关键词识别方法研究[D]. 北京:清华大学,1997.
- [16] 努尔麦麦提·尤鲁瓦斯,吾守尔·斯拉木. 面向大词汇量的维吾尔语连续语音识别研究[J]. 计算机工程与应用,2013,49(9):115-119.
- [17] 努尔麦麦提·尤鲁瓦斯,吾守尔·斯拉木,热依曼·吐尔逊. 维吾尔语连续语音识别声学模型优化研究[J]. 计算机工程与应用,2013,49(2):145-147.

(下转第305页)

保证的。

若伪造者想要伪造 $c = w \text{mod} q$, 他必须知道 w , 而 $w = e(P, P_2)^n R = nP$ 根据解离散对数的困难性, 已知 R , 无法得到 n 故无法得到 w , 因此无法伪造密文信息。对于 $c' = e(P, P_0) c \text{mod} q$ 其中 P_0 的值只有代理方知道, 而且 c 满足保密性, 伪造者无法得到任何相关信息。最后, 签名的过程中包含了授权者的私钥、密文信息, 对于密文, 该方案满足保密性, 所以攻击者是无法进行伪造的。

3.5 效率

在本文的效率分析中, 假设 G_1, G_2 分别表示加法和乘法运算的成本, 双线性对的计算成本为 e , 哈希函数的计算成本为 H , 指数运算的成本为 exp 。则通过与其它现有方案的效率比较, 得出如表 1 所示。

表 1 本文方案与其他方案的效率比较

方案	$G_1 / +$	G_2 / \times	e	H	exp
文献 [13]	$5G_1$	$6G_2$	$8e$	$7H$	$4exp$
文献 [14]	$8G_1$	$5G_2$	$9e$	$2H$	$5exp$
本文	$5G_1$	$4G_2$	$7e$	$4H$	$4exp$

由表可知, 本文在 G_1, G_2 和双线性对计算方面的成本相对其它方案均有所减少, 虽然其它计算成本并非最低, 但相对而言已达到很好的效果。另外, 转换密钥 P_0 的计算由代理方完成, 极大程度简化用户方的运算复杂性。代理方面计算 (U', c') 快速简单, 哈希过程与双线性计算成本大大减少。最后阶段通过验证等式 $e(V', U') = e(P_2 - P_1, P) \cdot e(P, P)^{H_2(c')}$ 的正确与否, 计算简便快捷, 减少了通信量和系统开销。

4 结 语

签密是将签名与加密有效结合。本文在传统计算的基础上, 通过加密、确认、代理重加密及解密等阶段, 对数据作一系列处理, 在确保消息正确、安全的前提下, 利用代理重加密能够转换密钥、二次加密的优势, 将信息再次加密, 被授权用户需要时可以用自己的私钥解密。增加用户向代理方确认阶段, 确保数据来源清楚, 使得消息发送者对于自己发出的消息不能抵赖。

代理重签密方面的相关方案日益成熟, 目前更多的应用于电子邮件转发、代理服务器等的使用中, 而安全问题作为信息传递中的主要障碍和制约因素, 关系到使用代理计算的相关企业的生存和发展, 因此还需要做更多的尝试与深入的研究。

参 考 文 献

[1] 雷万云. 云计算[M]. 北京: 清华大学出版社, 2011.
 [2] 张焕国, 王张宣. 密码学引论[M]. 武汉: 武汉大学出版社, 2009.
 [3] Rivest R L, Shamir A, Tauman Y. How to leak a secret[M]. Advances in Cryptology-ASIACRYPT 2001. Springer Berlin Heidelberg, 2001: 552-565.
 [4] Rivest R L, Shamir A, Adleman L. A method for obtaining digital signatures and public-key cryptosystems[J]. Communications of the ACM, 1978, 21(2): 120-126.
 [5] Bruce Schneier. Applied Cryptography: Protocols, Algorithms and Source Code in C[M]. 2nd ed. Wiley, 1995.
 [6] Blaze M, Bleumer G, Strauss M. Divertible protocols and atomic proxy cryptography[C]//Advances in Cryptology-EUROCRYPT'98. Springer Berlin Heidelberg, 1998: 127-144.
 [7] Ateniese G, Fu K, Green M, et al. Improved proxy re-encryption schemes with applications to secure distributed storage [J]. ACM

Transactions on Information and System Security (TISSEC), 2006, 9(1): 1-30.
 [8] Green M, Ateniese G. Identity-based proxy re-encryption[C]//Applied Cryptography and Network Security. Springer Berlin Heidelberg, 2007: 288-306.
 [9] Liang K, Liu Z, Tan X, et al. A CCA-Secure identity-based conditional proxy re-encryption without random oracles[C]//Information Security and Cryptology-ICISC 2012. Springer Berlin Heidelberg, 2013: 231-246.
 [10] Kawai Y, Takashima K. Fully-Anonymous Functional Proxy-Re-Encryption[J]. IACR Cryptology E-Print Archive, 2013: 318-391.
 [11] Gamage C, Leiwo J, Zheng U. An Efficient Scheme for Secure Message Transmission Using Proxy-signcryption[C]//Proc of the 22nd Australasian Computer Science. Auckland: Springer-Verlag, 1999: 420-431.
 [12] Zhang Xuejun, Wang Yumin. Efficient Identity-based Proxy Signcryption [J]. Computer Engineering and Applications, 2007: 43(3): 109-111.
 [13] 陈善学, 周淑贤, 姚小凤, 等. 高效的基于身份的代理签密方案[J]. 计算机应用研究, 2011, 28(7): 2694-2696.
 [14] 王会歌, 王彩芬, 曹浩, 等. 新的基于身份的代理重签密[J]. 计算机应用, 2011, 31(11): 2986-2989.

(上接第 223 页)

[18] 努尔麦麦提·尤鲁瓦斯, 吾守尔·斯拉木, 热依曼·吐尔逊. 基于音节的维吾尔语大词汇连续语音识别系统[J]. 清华大学学报: 自然科学版, 2013, 53(6): 741-744.
 [19] 努尔麦麦提·尤鲁瓦斯, 吾守尔·斯拉木, 热依曼·吐尔逊. 维吾尔语大词汇语音识别系统识别单元研究[J]. 北京大学学报: 自然科学版, 2014, 50(1): 149-152.
 [20] Takebayashi Y, Tsuboi H, Kanazawa. A Robust Speech Recognition System Using Word-Spotting with Noise Immunity Learning [C]//IC-AASP, 1991, 2(1): 905-908.

(上接第 254 页)

[13] An S J, Liu W Q, Venkatesh S. Face recognition using kernel ridge regression [C]//CVPR 2007: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, June 17-22, 2007, Piscataway, NJ: IEEE Press, USA, 2007.
 [14] Hoerl A E, Kennard R W. Ridge regression: Biased estimation for non-orthogonal problems[J]. Technometrics, 1970, 12(1): 55-67.

(上接第 271 页)

[15] Zhang L P, Yu H J, Hu S X. Optimal choice of parameters for particle swarm optimization [J]. Journal of Zhejiang University: Natural Science, 2005, 6(6): 528-534.
 [16] Janson S, Martin M. A hierarchical particle swarm optimization for dynamic optimization problems [C]//EvoWorkshops, 2004, LNCS 3005, 2004: 513-524.
 [17] 王俊伟, 汪定伟. 一种带有梯度加速的粒子群算法[J]. 控制与决策, 2004, 19(11): 1298-1300, 1304.
 [18] 肖健梅, 李军军, 王锡淮. 梯度微粒群优化算法及其收敛性分析[J]. 控制与决策, 2009, 24(4): 560-564.
 [19] 李丽娜, 曾庆勋, 甘晓晔, 等. 基于势函数与压缩感知的欠定盲源分离[J]. 计算机应用, 2014, 34(3): 658-662.
 [20] 何继爱, 何勇, 肖丹丹. 基于 KM-PCA 稀疏信号的盲源分离算法[J]. 兰州理工大学学报, 2012, 38(4): 80-84.
 [21] 董天宝, 杨景曙. 稀疏盲源分离快速算法[J]. 活力与指挥控制, 2012, 37(7): 84-87.