

基于正则表达式的语音识别控制策略研究

宋鑫坤¹, 陈万米^{1,2}, 朱 明¹, 桂春胜¹, 程硕远¹, 陈海波¹

(1. 上海大学 机电工程与自动化学院, 上海 200072;

2. 上海市电站自动化技术重点实验室, 上海 200072)

摘 要:文中提出一种新型控制策略,把正则表达式和服务机器人的语音控制结合起来,前者利用预定义的正则表达式对语音识别内容进行模式检出,后者根据检出的模式对机器人进行策略控制。通过分析识别文本构建正则表达式,使用微软 SAPI 构建客户端语音引擎进行语音识别,在服务机器人平台上验证策略的有效性、实用性和快速性。该新型控制策略把正则表达式应用在语音识别控制策略中,使语音交互技术将人类之间的交流方式应用于人类与机器之间,使机器人用正则表达式对人类语言进行词法分析实现自然语言的理解。提出的方法使机器人能够“理解”同一个语音命令的多种表达方式,使服务机器人的语音识别控制策略更具柔性和自然性,具有重要的应用价值和前景。

关键词:正则表达式;语音识别;服务机器人;控制策略

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2010)02-0106-04

Study on Speech Recognition Control Strategy Based on Regular Expression

SONG Xin-kun¹, CHEN Wan-mi^{1,2},

ZHU Ming¹, GUI Chun-sheng¹, CHENG Shuo-yuan¹, CHEN Hai-bo¹

(1. School of Electromechanical Engineering and Automation, Shanghai University, Shanghai 200072, China;

2. Shanghai Key Laboratory of Power Station Automation Technology, Shanghai 200072, China)

Abstract: A new control strategy combined regular expression and speech control system of service robot was proposed in the paper. The former implements pattern detection for the contents of speech recognition, and the later implements control strategies for the service robot based on the pattern detected. The regular expressions used to detect pattern was built through analyzing the source texts, the client voice engine used for speech recognition was built with Microsoft SAPI, the control strategies was examined on service robot platform and proved to be effective, practical and fast. The method proposed in this paper enables the robot to “understand” multiple expressions way for the same meaning, which makes the human-robot voice interactive more flexible and natural.

Key words: regular expression; speech recognition; service robot; control strategy

0 引 言

语音是最符合人类自然习惯的一种交互方式。随着语音交互技术的发展,它在人机接口和多媒体方面的应用也越来越普遍。而语言是人类用来交流最自然、最有效的手段,也是众多载体中具有最大信息量的信号,具有最高的智能水平。因此,语音交互和自然语言理解成为衡量机器智能的标准之一,在该领域的任何突破和新技术必将促进人工智能的发展和社会的进

步。

语音识别是服务机器人的必备技术,通常的语音识别内容局限于特定词或特定语句,不具备语言内容变化的适应性,而人们传递同一个意思往往使用多种表达方式。这就要求机器人必须具备自然语言理解的能力,采用词法分析器即可实现^[1]。关于词法分析有右线性文法、正则表达式、自动机等各种方法,它们的描述能力是等价的,但是复杂度各不相同。考虑到正则表达式简单优美的特质,系统采用正则表达式进行词法分析实现自然语言理解。

正则表达式经常应用于文本搜索和文本处理中,是实现模式匹配的强大工具^[2~4]。文中把正则表达式和服务机器人的语音识别控制策略结合起来,前者利用预定义的正则表达式对语音识别内容进行模式检

收稿日期:2009-05-30;修回日期:2009-08-07

基金项目:上海市重点学科建设项目(T0103)

作者简介:宋鑫坤(1984-),男,硕士研究生,研究方向为智能移动机器人、机器视觉;陈万米,副教授,研究方向为智能移动机器人、机器视觉、多机器人协调。

出,后者根据检出的模式对机器人进行策略控制。文中提出的方法使机器人能够“理解”同一个语音命令的多种表达方式,使服务机器人的语音识别控制策略更具柔性和自然性。

1 正则表达式

正则表达式(Regular Expression,简称 RE)是由数学家 Stephen Kleene 于 1956 年提出的。他的描述能力与 Chomsky 提出的三型文法,即右线性文法等价,主要通过字符的格式匹配进行词法分析,后来被推广为在静态文本中查找和替代的一种标准。

正则表达式是一个模式字符串,用于在给定字符串中搜索与之匹配的子串。虽然正则表达式书写困难,可读性也很差,但是正则表达式提供了功能强大、灵活而又高效的方法来处理文本,尤其是对于大量的未知文本提供了方便、快速的匹配方式。它的模式匹配法可以快速的分析大量的文本以找到特定的字符模式;并且可以提取、编辑、替换或删除字符串。正则表达式由普通字符、原义字符和特殊字符等元字符组成,能够按照特定语法规则被解释成多种字符串,并以此对目标字符串进行匹配。在实际应用中,正则表达式可以让用户通过使用一系列特殊字符构建匹配模式,然后把匹配模式与目标对象进行比较,根据比较对象中是否包含匹配模式来执行相应处理操作。由于正则表达式字符处理能力强大、灵活高效、适应性强,因此被广泛应用于字符串校验和分析处理不同领域的问题^[5-8]。

在服务机器人的语音识别系统中,识别内容往往是设定好的,就是说某个语音命令只对应仅有的一个或几个语音内容,严重限制了人机语音交流的柔性和自然性。从语义角度考虑,尽管说话内容各不相同,但它们是否表达了相同意思可以通过某个设定的模式进行有效地检测,一方面可以提取说话内容的寓意,另一方面也进行排除检测,避免误识别及误操作。正则表达式正是这样一种工具,Boost 库中的正则表达式能够实现对 Unicode 字符的支持,使 boost::regex 库能够对汉语文字进行语法分析和信息提取。

在正则表达式中,字母、数字和下划线都属于普通字符,许多标点符号具有特殊含义,这些特殊字符包括:~ \$. * + ? = ! : | \ / () [] { }。它的核心语法包括字符类、重复限定符、锚字符、子模式、贪婪匹配与非贪婪匹配模式。在文中应用到的特殊字符在表

1 中给出说明。

表 1 正则表达式语法

字符	匹 配
[...]	字符集操作,例如[abc]将匹配任何单个字符'a','b','c'
[^...]	字符集否操作,例如[^ a-c]表示 a 至 c 之外的所有字符
(...)	组合,将几个项目组合为一个元字符
?	匹配它前面的单元出现零次或者一次,等价于{0,1}
+	匹配它前面的单元出现任意次,等价于{0,1}
*	匹配它前面的单元至少出现 1 次,等价于{1,}
	或操作,例如 ab(d ef)匹配 abd 或则 abef
^	匹配目标字符串的开始
\$	匹配目标字符串的结尾

2 语音识别系统设计

语音引擎是在 Microsoft Speech SDK 的基础上独立开发,在经过适当的训练之后,非特定人语音识别率可以达到 90% 以上,能够实现简单的命令控制和语言理解。Microsoft Speech SDK 完全支持简体中文语音系统的开发,是开发语音软件的一个理想工具。它是基于 COM 标准开发,底层协议以 COM 组件的形式完全独立于应用程序层,为应用程序设计人员屏蔽了复杂的语音技术,充分体现了 COM 的优点。语音识别由识别引擎(Recognition Engine)管理,能实现从语音中识别出汉字;语音合成由语音合成引擎(也即为文本到语音引擎 TTS Engine)负责,实现语音朗读。程序员只需专注于自己的应用,调用相关的语音应用程序接口(SAPI)来实现语音功能,其结构如图 1 所示^[9]。

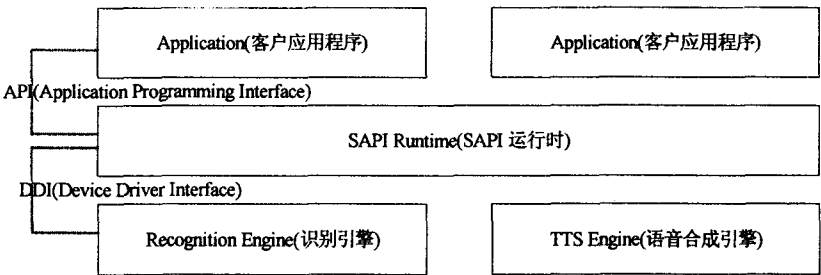


图 1 微软 Speech SDK 结构

为便于应用程序设计及模块调用,文中充分利用 C++ 语言的面向对象特性,将 SAPI 函数封装为 CSREngine 类,由 CSREngine 对象提供语音相关的所有接口包括语音初始化、语音设置、语音识别、语音合成、语音训练等功能。

该语音框架主要靠 SAPI 运行时(SAPI Runtime)实现应用程序与语音引擎之间的协作,而 SAPI 提供各种接口来实现不同的语音功能,如识别、朗读、训练等。与控制命令识别有关的接口及其方法主要包括:

(1)语音识别引擎(ISpRecognizer)接口:用于创建

语音识别引擎的实例。语音识别引擎对象有两种:独占(InProcRecognizer)的引擎和共享(SharedRecognizer)的引擎。独占的引擎对象只能由创建的应用程序使用,而共享的引擎可以供多个应用程序共同使用。

(2)语音识别上下文(ISpRecoContext)接口:主要用于发送和接收与语音识别相关的消息通知,创建语法规则对象。

(3)语法规则(ISpRecoGrammar)接口:定义引擎需要识别的具体内容,用于创建、载入和激活识别用的语法规则。而语法规则定义了期望识别的单词、短语和句子,通常有两种语法规则:听写语法(Dictation Grammar)和命令控制语法(Command and Control Grammar)。命令控制语法主要用于识别用户在语法文件里自定义的一些特定的命令词汇和句子,这些语法规则以 XML 文件的格式编写,通过(ISpRecoGrammar)接口载入,并激活。

(4)识别结果(ISpPhrase)接口:用于获取识别的结果,包括识别的文字,识别的语法规则等。

(5)语音合成(ISpVoice)接口:主要功能是通过访问 TTS 引擎实现文本到语音的转换。

语音识别的功能主要由 ISpRecognizer、ISpRecoContext、ISpRecoGrammar、ISpPhrase 四个接口实现,而且遵循 COM 组件的工作原理和 Windows 应用程序的工作原理(消息驱动机制),语音识别软件系统流程如图 2 所示,硬件只需在原来的机器人系统上配置麦克风。

软件系统实现分以下四个步骤:

- ①首先初始化 COM 平台;
- ②定义各个语音接口对象(以特定的顺序),设置识别语法、识别消息,使识别引擎处于工作状态;
- ③当有语法规则被识别后,ISpRecoContext 对象向应用程序发出语音识别消息,从而调用识别消息响

应函数;在消息响应函数中,通过 ISpPhrase 接口获取识别的结果。此步可以循环,直到停止语法规则为止;

④应用程序退出时,卸载 COM 平台。

3 正则表达式应用于语音识别控制策略

由于识别内容的多样性,并且人们习惯使用不同的表达方式传递同一个意思,所以,机器人能否从多样性的识别内容中或者多种表达方式中领悟到相同的意思就是一个非常具备挑战性的课题,即自然语言理解问题。这就要求机器人必须具备自然语言理解的能力,自然语言理解通常采用词法分析器实现。关于词法分析有右线性文法、正则表达式、自动机等各种方法,但是复杂度各不相同。考虑到正则表达式简单优美的特质,系统采用正则表达式进行词法分析实现自然语言理解。

基于正则表达式的自然语言理解是基于这样的共识:同一个意思通常可以用某一个固定模式来代表。这样的理论也在其它领域得到了应用和验证,如现代汉语语法处理、网络信息滤除、信息搜集等。

所以采用正则表达式实现模式检出进行简单的自然语言理解,从而实现同一个控制命令对应语言内容的多样性。在这里实现前进和导航功能,使机器人在接收到“前进/向前走/请前进/请向前走”时执行前进命令,接收到“导航到电视/导航到电视机/导航到 TV/请导航到电视/请导航到电视机/请导航到 TV”时执行导航命令。

3.1 编写语法规则文件(test.xml)定义待识别命令

语法规则文件定义了语音识别的范围,即预设词汇库,该文件以 xml 格式提供给主程序调用。文件内容如下。

```
<RULE ID="VID_Commands" TOPLEVEL="ACTIVE">
  <O>请</O>
```

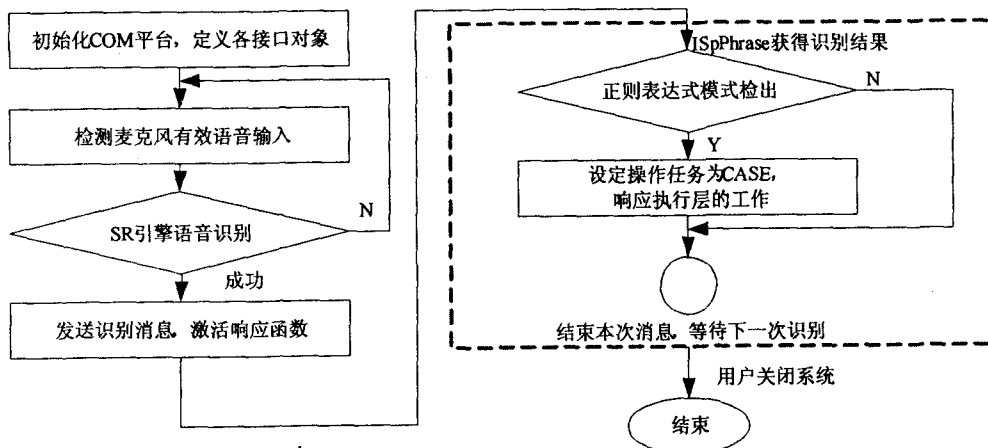


图 2 服务机器人语音识别软件系统设计

```
<O>导航到</O>
<P>
<RULEREF REFID="VID_CMD_Motion" />
</P>
</RULE>
<RULE ID="VID_CMD_Motion">
<L PROPID="VID_CMD_Motion">
<P VAL="VID_CMD_UP1">前进</P>
<P VAL="VID_CMD_UP2">向前走</P>
<P VAL="VID_CMD_TV1">电视</P>
<P VAL="VID_CMD_TV2">电视机</P>
<P VAL="VID_CMD_TV3">TV</P>
</L>
</RULE>
```

3.2 定义正则表达式

boost 库是一个经过千锤百炼、可移植、提供源代码的 C++ 库,作为标准库的后备,是 C++ 标准化进程的发动机之一。boost 库相当于 STL 的延续和扩充,它的设计理念和 STL 比较接近,都是利用泛型让复用达到最大化。不过对比 STL,boost 更加实用。STL 集中在算法部分,而 boost 包含了不少工具类,可以完成比较具体的工作。boost::regex 正则表达式库是其中核心的一部分,是用来做字符串分析及文本处理的重要工具。重要的是,boost::regex 提供了对各种编码的支持。在实际应用开发中,有几种常用的编码方案,包括针对中文编码的 GB2312、GB18030、UTF7、UTF8 等,还有统一编码方案 Unicode。<regex.hpp> 中有下面两行代码:typedef basic_regex<char> regex; typedef basic_regex<wchar_t> wregex;中文编码的正则表达式用 regex 实现就行了。文中用到的函数说明如下:boost::regex 用来生成和处理一个正则表达式;boost::cmatch 用来保存匹配的结果;boost::regex_search 用来搜索字符序列中哪些子序列是匹配输入的正则表达式的。文中定义正则表达式及声明匹配结果变量如下:

```
boost::regex RX_CMD_UP("请?(前进|向前走)");
boost::regex RX_CMD_TV("请?(导航到)?(电视|电视机|TV)");
boost::cmatch result;
```

3.3 识别内容模式检出

利用上一节介绍的 boost::regex_search 函数对识别结果进行模式检出,利用检出的模式实现机器人的控制策略转换。即机器人首先对识别内容进行词法分析,经过自然语言理解后再执行决策。控制策略如下:

```
void CDlgVoiceInteract::ExecuteCommand(char * recog_result)
```

```
{
    if ( boost::regex_search( recog_result, result, RX_CMD_UP))
    {
        m_player.decideMode(CMD_UP);
        goto TAG_N1;
    }
    if ( boost::regex_search( recog_result, result, RX_CMD_TV))
    {
        m_player.decideMode(CMD_TV);
        goto TAG_N1;
    }
TAG_N1:
    return;
}
```

为验证系统的有效性、实用性和快速性,文中对服务机器人设置三个正则表达式,分别进行 100 次识别实验,实验结果如表 2 所示。

表 2 正则表达式模式检出

正则表达式	语言内容变化形式	成功模式检出	平均耗时(ms)
RX_CMD_UP	4	98	500
RX_CMD_STOP	8	100	300
RX_CMD_TV	6	96	800

4 结束语

文中提出一种新型控制策略,把正则表达式和服务机器人的语音控制结合起来,把正则表达式应用在语音识别控制策略中,使语音控制策略将人类之间的交流方式应用于人类与机器之间,使机器人用正则表达式对人类语言进行词法分析实现自然语言的理解,是机器人智能水平的重要标志;并且与其它自然语言理解技术相比,这种实现方式简单有效,灵活多变,容易维护和定制。实验表明该方法能够较好地提高服务机器人语音识别控制策略的柔性和自然性,具有重要的应用价值和应用前景。

参考文献:

[1] 魏 蓉,王文忠,仲兰芬.正则表达式在现代汉语语法处理中的应用[J]. 阴山学刊,2007,21(4):48-52.

[2] Estrade B D,Perkins A L,Harris J M. Explicitly Parallel Regular Expressions[J]. Computer and Computational Sciences, 2006,1(1):402-409.

[3] Arslan A N. Multiple Sequence Alignment Containing a Sequence of Regular Expressions[J]. Computational Intelligence in Bioinformatics and Computational Biology,2005,14(15):1-7.

类样本的规模增大,分类的性能稳步上升,分别从77.03、72.73等上升到80.88、77.12等。最终完成了对未标样本的分类。

为了验证在基于小规模样本环境下主动学习的优势,以经典的受监督学习方法贝叶斯分类器进行分类。分别以某一网站的50%数据作为训练集,剩下的数据与其他网站的数据作为分类样本。实验结果见表4。

表4 简单分类器下的分类性能

	163	sina	Digitalchina	pconline	jwdigital
F 值	42.06	34.96	42.12	52.20	44.87

由于训练样本数据与分类样本具有交大的差异,分类样本中不确定性样本的比例较大,导致学习器在对未知数据进行分类时分类性能不理想。这说明训练样本数据的选择对于未标分类样本的分类有很大的影响。

针对这一问题,采用文中的信息熵主动学习方法,通过引入信息熵的信息量化度理论,每次迭代选取最大适应于当前分类器的样本进行分类同时完善训练数据,优化分类器的参数。有效利用了未标样本内部信息,提高了分类器的性能。

4 结束语

主要针对小规模训练样本环境下,如何对含有大量潜在不确定因素的未标样本集进行分类。通过引入主动学习的策略,基于信息熵的信息量化理论,利用未标样本的内在信息进行学习分类。实验结果表明,该方法具有较好的分类性能。

参考文献:

- [1] Rocchio J. Relevant feedback in information retrieval[M]//In Salton G. The smart retrieval system - experiments in automatic document processing. Englewood Cliffs, NJ: [s. n.], 1971.
- [2] McCallum A, Nigam K. A comparison of event models for naive Bayes text classification[C]//AAAI - 98 Workshop on Learning for Text Categorization. [s. l.]: AAAI Press, 1998.
- [3] Guyon I, Boser B, Vapnik V. Automatic capacity tuning of very large Vcdimension classifiers[J]. Advances in Neural Information Processing Systems, 1993(5): 147 - 155.
- [4] Igam K, McCallum A, Thrun S, et al. Learning to classify text from labeled and unlabeled documents[C]//In: Mostow J, Madison C R. Proceedings of the 15th National Conference on Artificial Intelligence. Wisconsin: AAAI Press, 1998: 792 - 799.
- [5] 刘 晶, 郭 雷, 聂晶鑫. 基于 SVM 的一种新的分类器设计方法[J]. 计算机应用研究, 2006, 23(7): 181 - 183.
- [6] Engelbrecht A P, Cloete I. Incremental Learning Using Sensitivity Analysis[C]// Neural Networks, 1999. IJCNN apos; 99. International Joint Conference. [s. l.]: IEEE Press, 1999: 1350 - 1355.
- [7] 陈耀东, 王 挺, 陈火旺. 半监督学习和主动学习相结合的浅层语义分析[J]. 中文学习学报, 2008, 22(2): 70 - 75.
- [8] Thompson C A, Califf M E, Mooney R J. Active Learning for Natural Language Parsing and Information Extraction[C]//In: Proceedings of the sixteenth International Machine Learning Conference. Slovenia: [s. n.], 1999.
- [9] 张健沛, 徐 华. 支持向量机主动学习方法研究与应用[J]. 计算机应用, 2004, 24(3): 1 - 3.
- [10] Cohn D A, Ghahramani Z, Jordan M I. Active learning with statistical models[J]. J. of Artificial Intelligence Research, 1996, 4: 129 - 145.
- [11] Liere R, Tadepalli P. Active learning with committees for text categorization[C]//In Proceedings of the Fourteenth National Conference on Artificial Intelligence. Providence, RI: [s. n.], 1997: 591 - 596.
- [12] McCallum A, Nigam K. Employing EM and pool - based active learning for text classification[C]//In Machine Learning: Proceedings of the Fifteenth International Conference (ICML '98). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, 1998: 359 - 367.
- [13] Carter T. An introduction to information theory and entropy [EB/OL]. 2007 - 06. <http://astarte.csustan.edu/tom/SFICSSS>.
- [4] Brodie B C, Taylor D E, Cytron R K. A Scalable Architecture For High - Throughput Regular - Expression Pattern Matching[J]. Computer Architecture, 2006, 11(1): 191 - 202.
- [5] 陆 虎, 宋余庆, 薛万宇, 等. 一种基于正则表达式匹配的协议分析异常检测方法[J]. 计算机应用与软件, 2008, 25(3): 118 - 122.
- [6] 李哲夫. 正则表达式在电信业务处理中的应用研究[M]. 广州: 暨南大学出版社, 2008.
- [7] 余石泉, 周肆清. 正则表达式在编程题自动阅卷中的应用[J]. 计算机技术与发展, 2007, 17(7): 109 - 112.
- [8] 王功明, 吴华瑞, 赵春江, 等. 正则表达式在电子政务客户端校验中的应用[J]. 计算机工程, 2007, 33(9): 140 - 143.
- [9] Singh C T, Maulik U. A framework for an artificial immunity and speech based navigation for mobile robots[J]. Evolutionary Computation, 2008, 3(4): 1247 - 1251.

(上接第 109 页)