

语音识别综述

崔文迪, 黄关维

(厦门大学计算机系 福建 厦门 361001)

【摘要】 语音作为一个交叉学科,具有深远的研究价值,近 50 年的研究发展,语音识别技术已经有了极大的发展,但大多数产品能存在与实验室,没有达到使用化的效果,所以语音识别的研究还要更加深入。本文介绍了语音识别的发展过程,以及一个语音系统框架和识别过程,HMM 模型的概念和建立,还有语音发展的问题和解决方案。

【关键字】 语音识别, 语音识别系统, HMM

1. 引言

让计算机能听懂人类的语言,是人类自计算机诞生以来梦寐以求的想法。随着计算机越来越向便携化方向发展,随着计算环境的日趋复杂化,人们越来越迫切要求摆脱键盘的束缚而代之以语音输入这样便于使用的、自然的、人性化的输入方式。尤其是汉语,它的汉字输入一直是计算机应用普及的障碍,因此,利用汉语语音进行人机交互是一个极其重要的研究课题。它正在直接与办公、交通、金融、公安、商业、旅游等行业的语音咨询与管理,工业生产部门的语音控制,电话、电信系统的自动拨号、辅助控制与查询,以及医疗卫生和福利事业的生活支援系统等各种实际应用领域相接轨,并且有望成为下一代操作系统和应用程序的用户界面了。[1]

2. 语音识别系统构成[2]

语音系统基本构造如图 1 所示,系统可以分为前端处理和后端处理,前端处理包括语音的录入、处理、特征值的提取,后端是个数据库的搜索过程,分为训练和识别,训练是对所建的模型进行评估、匹配、优化,获得模型参数,识别是一个专用的搜索数据库,获取前端数值后,在声学模型、一个语言模型和一个字典,声学模型表示一种语言的发音声音,可以通过训练来识别特定用户的语音模型和发音环境的特征,语言模型是对语料库单词规则化的概率模型。字典列出了大量的单词及发音规则。总体上说,语音识别是一个模式识别匹配的过程。在这个过程中,计算机首先要根据人的语音特点建立语音模型,对输入的语音信号进行分析,并抽取所需的特征,在此基础上建立语音识别所需的模板。然后,在识别过程中,计算机根据语音识别的整体模型,将计算机中已经存有的语音模板与输入语音信号的特征进行比较,并根据一定的搜索和匹配策略找出一系列最优的与输入语音匹配的模板。最后通过查表和判决算法给出识别结果。显然,识别结果与语音特征的选择、语音模型和语言模型的好坏、模板是否准确等都有直接的关系。



图1 语音识别系统

3. 语音的特征提取

特征提取就是要从语音波形中提取出重要的反映语音特征的相关信息,而去掉那些相对无关的信息,如背景噪声、信道失真等,并把这些信息转换为一组离散的特征参数。目前常用的特征参数有下面两种。

3.1 线性预测倒谱技术(LPCC)

线性预测分析技术是目前广泛被使用的特征参数提取技术,特别是在早期的许多成功的系统中很多都是用 LPC 倒谱系数作为系统的特征矢量。

这里的 LPC 倒谱实质上是复倒谱,复倒谱是指信号通过 Z

变换以后去对数,再求反 Z 变换而得到。线性预测分析的声道模型系统函数 $H(z)$ 反映了声道的频率响应和原始信号的谱包络,因此用 $\lg H(z)$ 作反 Z 变换即可求出其复倒谱系数(LPCC)。LPCC 的优点在于计算量小,易于实现,缺点在于抗噪性能差,在对于汉语识别的时候对辅音识别差,所以一般我们采取的是 Mel 频率倒谱系数(MFCC)。

3.2 Mel 频率倒谱系数(MFCC)

目前大部分的语音识别的特征提取都采用 Mel 频率倒谱系数,这是因为 Mel 刻度在对声学测量时是最合理的频率刻度。MFCC 参数具有良好的识别性能和抗噪能力,但其计算量和精度都要求很高。Mel 刻度与频率的转换关系为

$$F_{\text{mel}} = 3322.23 \lg(1 + 0.001 f_{\text{Hz}}) \quad (1)$$

4. 语音模型的选择

4.1 HMM 模型

HMM 模型是语音信号时变特征的有参表示法。它由相互关联的两个随机过程共同描述信号的统计特性,其中一个隐藏的(不可观测的)具有有限状态的马尔可夫链,另一个是与马尔可夫链的每一状态相关联的观察矢量的随机过程(可观测的)。隐马尔可夫链的特征要靠可观测到的信号特征揭示。这样,语音等时变信号某一段的特征就由对应状态观察符号的随机过程描述,而信号随时间的变化由隐马尔可夫链的转移概率描述。

HMM 模型在某状态 j 下对应的观察值可以由一组概率 b_{jk} , $k=1,2,\dots,M$ 来描述,它是个 M 个离散可数的观察值中的一个,因而称为离散 HMM。当观察值为一个连续的随机变量 X ,其在状态 j 下对应的观察值由一个观察概率密度函数 $b_j(X)$ 表示,这就成了连续的 HMM。连续的 HMM 用 Baum-Welch 算法估计模型参数的时,虽然在估计 π, A 参数时适用,但在估计描述 $b_j(X)$ 的参数时必须对 $b_j(X)$ 加以一定的限制才能成立。目前运用最广泛的是高斯型 $b_j(X)$ [9]它可以用下面公式表示:

$$b_j(X) = \sum_{k=1}^K c_{jk} b_{jk}(X) = \sum_{k=1}^K c_{jk} N(X, \mu_{jk}, \Sigma_{jk}) \quad (2)$$

$1 \leq j \leq N$

其中, $N(X, \mu_{jk}, \Sigma_{jk})$ 为多维高斯概率函数, μ_{jk} 为均值矢量, Σ_{jk} 为方差矩阵, K 为 $b_j(X)$ 的混合概率个数, $c_{jk}(X)$ 为组合系数,且

$$\sum_{k=1}^K c_{jk} = 1 \quad (3)$$

4.2 HMM 中的 3 个基本问题及其解决方案

欲使所建立的 HMM 模型能够解决实际问题,以下 3 个问题必须加以解决:

1) 已知观察序列 O 和模型 $\lambda=(A,B,\pi)$, 如何计算由此产生此观察序列概率 $P(O|\lambda)$?

这个问题实际上是一个模型评估问题,因为 $P(O|\lambda)$ 反映了观察序列与模型吻合的程度。在语音识别中,我们可以通过计算、比较 $P(O|\lambda)$, 从多个模型参数中选择出与观察序列匹配的最好的模型。为了解决这个问题,前人已经研究了向前向后算法。

2) 已知观察序列 O 和模型, 如何确认一个合理的状态序列,使之能最佳地产生 O , 即如何选择最佳的状态序列 $q=\{q_1, q_2, \dots$

q_i ?

这个问题关键是怎样最佳的准则来决定状态的转移。一种可能的最佳准则是:

$$|qt^* = \arg \max_i [P((1=i)) | O|\lambda] \quad (6)$$

这里存在一个问题: 有时候会出现不允许的转移, 即 $a_i = 0$, 那么对这些 i 和 j 所得到的状态序列就是不可能的状态序列也就是说, 式(13)得到的解只是在每个时刻决定一个最可能的状态, 而没考虑整体结构, 相邻的状态和观察序列长度问题。针对这个问题, 最好的解决方案是 Viterbi 算法, 也是在语音识别过程中的主要算法。

3) 语音模型训练的好坏直接关系到语音识别系统识别率的高低。为了得到一个模板, 往往需要有大量的原始语音数据来训练语音模型。因此, 在开始进行语音识别研究之前, 首先要建立起一个庞大的语音数据库和语料数据库。一个好的语音数据库包括足够数量、具有不同性别、年龄、口音说话人的声音, 并且必须要有代表性, 能均衡地反映实际使用情况。有了语音数据库及语音特征, 就可以建立语音模型, 并用语音数据库中的语音来训练这个语音模型。训练过程是指选择系统的某种最佳状态不断地调整参数 (A, B, π) , 使得 $P(O|\lambda)$ 最大。这是一个复杂的过程, 因为没有解析法可以用来求最大似然模型, 所以只能用迭代法 (Baum-Welch) 算法或者使用最佳梯度法。要求计算机有强大的计算能力, 并有很强的理论指导, 才能保证得到良好的训练结果。

5. 语言模型

语言模型对中、大词汇量的语音识别系统特别重要。当分类发生错误时, 可以根据语言学模型、语法结构、语义学进行判断纠正, 特别是一些同音字则必须通过上下文结构才能确定词义。语言学理论包括发音学、音韵学、语义结构、语言的数学描述模型等。把语言模型应用于语音识别中要解决两个问题: 一是能够用数学模型来描述语言中词的语言结构; 二是在给定这样一种结构的基础上, 如何把它和模式识别器结合找出一种有效的识别算法。目前比较成功的语言模型通常是采用统计语法的语言模型, 如二元文法 (Bigram) 和三元文法 (Trigram)

6. 语音识别面临的困难和解决方案

语音识别面临的主要困难是理论上没有突破。虽然各种新的修正方法不断涌现, 但其普遍适用性都值得商榷。此外, 语音识

(上接第 18 页)

<分类结果>:=蔷薇科 | 李亚科 | 李属 | 李...

3.3 专家系统的设计

专家系统的建立一般有知识获取、知识表达的选择、专家系统的建立三个步骤。一般情况下, 植物分类方面的专家不是开发专家系统的计算机工程师, 所以在知识获取这一环节中, 工程师与专家间一定要事先沟通, 使工程师存储在计算机中的知识能正确表达出专家的意见, 这是建立一个成功专家系统的前提。在知识表达选择中, 可根据专家系统设计的方法进行。

在基于产生式规则的专家系统中, 有一个用来构成系统的知识库形成规则集, 在植物分类专家系统中, 可以为每一种植物建立一个产生式规则, 规则头是植物名称的断言, 而规则则是体现这种植物所具有的特征。

在知识库的设计中, 将各种知识以具体植物种类为单位整合成一个模块, 如对植物分类这一对象来说, 可将相关的判断性、经验性、控制性等知识装入某一模块中, 则知识库中就有许多相对应的模块。同时, 在专家知识或用户需求发生变化时, 只需对某一模块进行修改即可。

在植物分类专家系统的推理机中有一个解释程序, 用以选择和激活各模块, 解释程序的工作可分为三步, 首先用知识库中的数据元素和规则模式进行分配; 如果可被激活的规则等大多于一个, 则运用冲突消解策略选择一个规则; 再利用选定的规则找出植物分类的结果。

3.4 专家系统的实现

植物分类专家系统的实现, 依赖于计算机一定的前台语言和后台系统的支持。程序设计语言是开发专家系统的最基本工

具。其中, Lisp 和 Prolog 这两种人工智能语言常用来开发专家系统, VB、DEPHI 等也是构造专家系统的常用语言。而后台的操作平台在现行的 Window 2000、XP 系统中均可实现。

1) 语音对环境的依赖

在某些环境下采集的语音进行训练后, 必须在相同的环境下进行识别, 否则性能急剧下降。还有麦克风质量的不同, 位置摆放的问题, 会导致用户输入的语音夹杂着噪音还有回音, 这样也会影响识别的准确性, 目前可采用麦克风阵列技术消除单一麦克风对语音的影响[11], 在预处理阶段通过语音增强算法, 使语音的懂度和信噪比增强。

2) 模型的改进和算法的更新

在识别算法上 HMM 算法得到进一步的发展, 提出了各种各样的 HMM 改进算法。加入了遗传算法, 还有神经网络等新技术[12], 使得 HMM 的训练和识别更加准确, 随着计算机的硬件发展, 并行算法与语音识别也可以进行结合, 采用并行技术[13] 使得语音系统对语料库的训练, 还有更快的进行搜索都奠定了基础。

参考文献:

1. 赵力. 语音信号处理. 第一版, 北京: 机械工业出版社, 2003
2. 王炳锡. 实用语音识别基础, 北京: 国防工业出版社, 2005
3. Young S. HTKHistory[EB/OL]. 2005. <http://htk.eng.cam.ac.uk>
4. Kai - Fu Lee. Automatic Speech Recognition. The Development of the SPHINX System
5. Reynolds D.A. and Rose R.C., Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans. Speech Audio Processing, Vol. 3, No. 1, Jan. 1995, pp. 72-83.
6. Chao Huang, Tao Chen and Eric Chang. SPEAKER SELECTION TRAINING FOR LARGE VOCABULARY CONTINUOUS SPEECH RECOGNITION.
7. Gu H. Isolated Utterance speech recognition using hidden Markov models with bounded states duration. IEEE Trans. SP, 1991, 39(8):1743-1752
8. 李净, 郑方. 汉语连续语言内识别中上下文相关声韵母建模. 清华大学学报(自然科学版) 2004 年, 第 44 卷, 第一期: 61-64
9. L. A. Liporace. Maximum Likelihood for Multivariate Observation of Markov Sources. IEEE Trans. IT, 1982, 28(5): 729-734

具。其中, Lisp 和 Prolog 这两种人工智能语言常用来开发专家系统, VB、DEPHI 等也是构造专家系统的常用语言。而后台的操作平台在现行的 Window 2000、XP 系统中均可实现。

在植物分类专家系统中, 产生式规则的数目决定了知识库的规模, 植物种类繁多, 系统初始实现时可先对一个地区、一个省份的植物种类进行设计, 然后再逐步扩展到更大范围, 在知识库中添加更多的规则, 使得植物分类专家系统得以扩展完善。

4. 小结

本文首先对植物的分类方法和专家系统进行了详细阐述, 设计了植物分类的层次模型, 提出了基于产生式规则的植物分类专家系统的设计方案。该系统的采用将使得植物分类的结果更加准确、高效, 具有一定的工程应用价值。

参考文献:

1. 胡伟, 李霞等. 基于专家系统和神经网络集成的植物智能分类系统[J]. 宜宾学院学报, 2005, 12: 69-72.
2. 罗燕琪, 陈雷霆. 专家系统中知识表示方法研究[J]. 电子计算机, 2001, 4: 25-31.
3. 王正军, 程家安等. 专家系统及其在害虫综合治理中的应用[J]. 江西农业学报, 2000, 12(1): 52-57.
4. 杨珺. 基于虚拟植物生长模型的分布式农业专家系统研究[J]. 微电子学与计算机, 2006, 23(A): 229-230.
5. 蔡自兴, 徐光佑. 人工智能及其应用. 北京: 清华大学出版社, 2003
6. 张丈量, 纪有奎. 专家系统原理与设计. 武汉: 武汉测绘科技大学出版社, 1989