

维吾尔语广播新闻连续语音敏感词检索系统^①

木合塔尔·沙地克^{1,2}, 李 晓¹, 布合力齐姑丽·瓦斯力³

¹(中国科学院 新疆理化技术研究所, 乌鲁木齐 830011)

²(中国科学院 研究生院, 北京 100084)

³(新疆教育学院 数学与信息技术分院, 乌鲁木齐 830043)

摘 要: 首先介绍语音信号来源于新疆人民广播电台维吾尔语新闻的敏感词语音语料库的建设。然后用该语料库进行基于 HMM 的模型训练。模型训练中详细介绍识别基元端点检测、特征提取、矢量量化、码本构建、HMM 模型训练过程和结果。最后用该语料库和 HMM 训练模型对维吾尔语广播新闻连续语音信号进行敏感词检索, 并对检索结果进行分析。

关键词: 语音语料库; 敏感词检索; 维吾尔语; 单词分割; 连续语音识别

Uyghur Broadcast News Continues Speech Sensitive-Word Spotting System

MU He Ta Rr Sha di ke^{1,2}, LI Xiao¹, BU He Li Qi Gu Li Wa Si Li³

¹(Xinjiang Technical Institute of Physics & Chemistry, Chinese Academy of Sciences, Urumchi 830011, China)

²(Graduate School, Chinese Academy of Sciences, Beijing 100084, China)

³(College of Mathematic and Information Technology, Xinjiang Education Institute, Urumchi 830043, China)

Abstract: First, this paper introduces the design a corpus based on Uighur news of Xinjiang Broadcast Station. And then, Training the HMM for this corpus. In this step, introduces Word Segmentation, Feature Extraction, Vector Quantization, Codebook, HMM Training etc. Finally, use the corpus and training model implements HMM based Uyghur broadcast news continues speech Sensitive-word spotting and give the conclusion of the test.

Key words: corpus; sensitive-word Spotting; Uyghur; word Segmentation; continues speech recognition

1 引言

1.1 维吾尔语简介

维吾尔语属阿尔泰语系突厥语族, 在结构语法上属于粘着语类型。维吾尔语除了固有词语和突厥语同源词之外, 借词在词汇中占有一定的比例。借词主要来自阿拉伯语、波斯语、俄语、汉语、蒙古语等。现代维吾尔语在语音上有和谐、弱化、脱落、增音等语音变化^[1]。

现行维吾尔文是阿拉伯文字为基础的拼音文字, 字母共有 32 个, 其中 8 个元音字母, 24 个辅音字母。元音发音时, 声带振动, 气流通过口腔时不受阻碍。元音均是浊音。发辅音时, 气流通过口腔时受到某种

阻碍。发音时按声带振动与否, 辅音可分为浊辅音和清辅音。浊辅音(发音时声带振动): b, j, d, r, z, zj, gh, g, ng, l, m, n, w, y。清辅音(发音时声带不振动): p, t, ch, x, s, sh, f, q, k, h。

在维吾尔语里词可由一个或几个音节构成, 而音节则可由一个或几个音组成。元音可以单独构成音节, 而辅音不能单独构成音节。元音是维吾尔语音节的标志, 一个词里有几个元音, 它就有几个音节(一些复合元音的汉语借词除外)。维吾尔语音节的构成方式有: V, VC, CV, CVC, VCC, CVCC 等六种。为了正确拼写借自汉语、俄语等语言的外来词, 自治区语委会在 1985 年修订维吾尔语标准语正字法时, 在书面语中

① 收稿时间:2011-07-08;收到修改稿时间:2011-08-07

增加了以下五种音节模式: CCV, CCVC, CCVCC, CVV, CVVC。

1.2 维吾尔语语音识别现状

维吾尔语语音识别研究工作开始于20世纪90年代初。1994年,吾守尔·斯拉木采用独特的音节训练词识别方法和词汇扩充方法等技术,研制出1200个识别音节、4万个识别词汇的联想式特定人维吾尔语语音识别系统,其识别率达到95%^[2]。2005年专门成立多语种语音信息处理研究室,开展了维吾尔语语音识别、语音合成的研发工作。

近年来,新疆大学、中国科学院新疆理化技术研究所、新疆师范大学等高校和科研机构,先后进行了很多维吾尔语语音识别与合成方面的技术研究,积累了一定的经验。然而,这些研究都是在高质量的实验条件下,采用标准发音的、仔细朗读的语音进行的。但是要真正完成一个能够处理自然语言识别系统,就现有的技术水平来说还是非常困难的。

2 语料库准备

语音语料库按文件组织,每个文件需有完整描述该文件所含语音的录音特性、说话人特性、发音特性以及数据特性等内容^[3]。

(1) 录音特性:录音时间,录音地点,录音通道,噪声环境等;其中录音时间用年月日表示,录音时间相差不超过3月的当作同一时间;录音通道包括PC、电话、嵌入等;具体指标描述应包括话筒类型、名称、灵敏度以及阻抗,通道的低频、高频截至频率。

(2) 说话人特性:姓名,性别,籍贯,文化程度,年龄,口音等方言区。

(3) 发音特性:包括孤立语音、孤立词语音、朗读连续语音、口语语音等。

(4) 数据特性:采样频率,采样数据点位数,通道数目,通道噪声,数据编码格式。

3 语音预处理

3.1 识别单元的选取

语音识别基元的选择在语音识别尤其是连续语音识别中是非常重要的环节^[3]。识别基元的选取应该基于如下几个原则:首先,对于不同的语言,基元选择的考虑是不同的,因为各种语言的结构是不同的。例如,对于维吾尔语,可以用音素,也可以用词等作为

识别单元。对于汉语,则可用“声母—韵母”,也可用音节字、词等作为识别基元。第二,识别基元选取具有灵活性,用它可以组成其他的语音或语法单位;具有稳定性,它应该在不同的语音环境和语言环境中相对稳定。灵活性希望基元尽可能地小,如音素;而稳定性则希望基元尽可能地大,如词甚至词组。然而,基元的选择还需考虑搜索时间和模型存储开销。基元选得愈小、存储量愈小,算法愈复杂,正识率也愈低。另外,基元选择也与实际用途有关。一般地说,有限词汇量的识别基元可以选得大一些(如:词或短语等);而无限词汇量的识别基元则应选得小一些(如:音素、声母—韵母等)。

3.2 自动分段-端点检测

端点检测是指用数字处理技术来找出语音信号中的各种段落(如:音素、音节、词素、词...等)的始点和终点的位置。在维吾尔语中,主要目的是找出音素或词的两个端点。这种处理对于维吾尔语连续语音识别是非常必要的。

在实际应用中,由于环境噪声的影响,给语音识别系统的稳定性带来严重的问题,使得语音识别系统的识别性能降低而不能实用。噪声语音信号的语音端点检测,是语音处理系统中极其关键的一步。

语音和噪声的主要区别在它们的能量上,语音段的能量比噪声段的大。要准确找出语音段的起止点,在短时能量检测方法的基础上,加上短时平均过零率(过零率表示一帧信号波形穿过横轴的次数,它实际上反映了信号的频率特征),利用能量和过零率作为特征来进行检测。这种方法也被称为双门限比较法。

在双门限比较法中,短时平均过零率定义为:

$$Z_n = \frac{1}{2N} \sum_{m=-\infty}^{+\infty} |sgn[x(m)] - sgn[x(m-1)]| w(n-m) \quad (1)$$

式(1)中,sgn[·]是符号函数,其定义为:

$$sgn[x] = \begin{cases} 1 & (x \geq 0) \\ -1 & (x < 0) \end{cases} \quad (2)$$

w(n)是窗函数,在这里一般用矩形窗,x(m)为原始语音信号取样序列,Zn的下脚注n是指窗的位置。

4 语音特征参数—LPCC

语音信号是十分复杂的非平稳信号,它不仅包含语义信息,还有个人特征信息,对其特征参数的研究

是语音识别的基础。特征参数应能完全、准确地表达语音信号所携带的全部信息。

线性预测倒谱参数(Linear Prediction Cepstrum Coefficient, LPCC)是线性预测系数 (Linear Prediction Coefficient, LPC) 在倒谱域中的表示。该特征是基于语音信号为自回归信号的假设, 利用线性预测分析获得倒谱系数。LPCC 参数的优点是计算量小, 易于实现, 对元音有较好的描述能力, 其缺点在于对辅音的描述能力较差, 抗噪声性能较差。

由于通过自相关法求得的 LPC 系数保证了系统的稳定性, 使得声道模型传输函数具有最小相位。利用这一特性, 可以推导出语音信号的 LPCC 与 LPC 系数之间的关系。是由 LPC 得到:

$$LPCC_n = LPC_n + \sum_{k=1}^{n-1} \frac{n-k}{n} LPCC_{n-k} LPC_k \quad (3)$$

我们知道, 语音信号所携带的语义信息主要体现在声道传输函数上, 因而在语音识别中通常取语音信号倒谱的低时域构成 LPCC, 即:

$$LPCC = [LPCC_1, LPCC_2, \dots, LPCC_q], 10 \leq q \leq 16 \quad (4)$$

其中, q 为 LPCC 特征的阶数。

5 语音识别模型—HMM

模型是对信号本质描述的数学方法之一。对于确定性信号用函数关系描述, 对于平稳随机信号用数理统计的方法描述, 而语音信号是非平稳随机复杂信号。研究人员根据任务要求, 创造性地建立了多种语音识别模型。模型的本质应属于过程逼近, 从识别的角度看, 模型应理解为分类模式。所有的模型都要训练才能用于识别, 而训练数据的选择方式及多少, 往往关系到模型的成败^[4]。

隐马尔可夫过程是一个双重随机过程: 一重用于描述非平稳信号的短时平稳段的统计特征 (信号的瞬态特征, 可直接观测到); 另一重随机过程描述了每个短时平稳段是如何转变到下一个短时平稳段, 即短时统计特征的动态特性 (隐含在观察序列中)。

人的语言过程也是这样一种双重随机过程。因为语音信号本身是一个可观察的序列, 而它又是由大脑里的 (不可观察的)、根据言语需要和语法知识 (状态选择) 所发出的音素 (词、句) 的参数流。大量试验证明, HMM 的确可以非常精确地描述语音信号的产

生过程。隐马尔可夫模型的数学描述请看文献[3]。

6 系统基本原理

维吾尔语广播新闻连续语音敏感词检索 (下文简称: 敏感词检索) 与连续语音识别有着密切的关系。首先, 敏感词检索的任务是在连续的无限制的自然的维吾尔语广播新闻连续语音流中检索出给定的敏感词, 因而它首先是一种连续语音识别。其次, 它无法对发出自然语音流的播音人提出任何要求, 因此敏感词检索又不同于连续语音识别。

敏感词检索系统因为其任务的特殊性, 系统组成与连续语音识别有一定区别。本系统的一般结构可用图 1 表示。

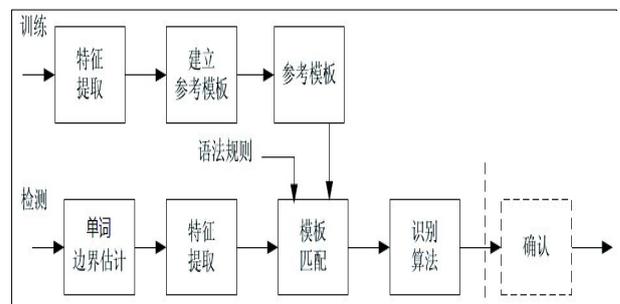


图 1 本系统一般结构图

该模型, 首先对输入语音进行粗略分段, 利用语音的能量和过零率大致分出词素边界。系统的识别性能很大程度上取决于前端词素边界估计的结果。因为敏感词检索系统不需要识别出所有发音, 除敏感词以外的都称为填料或废料, 所以敏感词检测系统的参考模板采用敏感词模板+填料模板的形式, 其语法网络结构如图 2 所示。其中, 系统的操作点可以通过设置词间的转移权重 $w_{kn}(1 \leq n \leq N)$ 和 $w_{fn}(1 \leq n \leq M)$ 来调节。

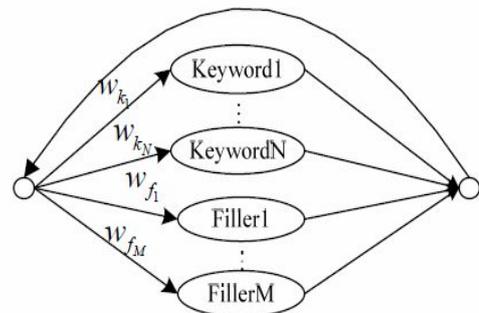


图 2 敏感词+填料的语法网络结构图

我们采用词素（本系统的词素：也可能是一个单词，也可能是单词的一个分段）为识别基元，任务是给定的维吾尔语广播新闻语音段中是否出现敏感词。若语音段中出现敏感词，显示第一个敏感词所出现的位置。

7 系统实现

7.1 语料库准备

虽然目前获得第一手语料较容易，但语料的手工标注等方面花钱费时困难大。本系统中，我们手工准备语音语料库。语料库的语音来源于新疆电台维吾尔语广播新闻，系统前端语音的采样率为 8000Hz，采样数据点位数 16 位，单声道。

本敏感词语料库含 3 个敏感词。每个敏感词从维吾尔语广播新闻连续语音不同语音段截取两次（即每个敏感词重复发音两次），每个发音的语音长度不一。语料库以数据结构形式定义的，有三个域（Info, Label 和 Case）。Info 域是字符串，存放如下内容：“时间：2011 年 5 月 30 日；地点：新疆人民广播电台维吾尔语频道新闻 60 分；通道：电台网站；姓名：阿不都瓦伊提·伊米提；性别：男；孤立词语音；采样率：8000Hz；采样数据点位数：16；通道数目：1；数据编码格式：wav”。Label 和 Case 域均是基元数组，域的长度表示数据结构包含的敏感词数目。Label 基元数组中存储每个敏感词文本字符串，而 Case 基元数组的每个基元又是的一个基元数组，其中每个基元存储每个敏感词的不同发音文件名。

7.2 识别基元端点检测

语音可分为浊音和清音。浊音由基本频率低于几百 Hz 的准周期的声乐线生成。清音由湍流气流的随机声音生成。浊音的幅度比清音的幅度大于十倍左右。另外，浊音的平均能量比清音的平均能量大得多。清音的过零率比浊音的过零率多得多。

我们用语音的短时能量和过零率来初步判断识别基元的端点。对给定的语音信号来说，短时能量和过零率用一定长度的（帧长）滑动窗口来计算出来。如果，短时能量大于已定的阈值并短时过零率小于已定的阈值则该滑动窗口的中间采样值被确定为浊音。能量阈值相对短时能量的最大值来表示。

图 3 是，短时能量阈值设置 0.1，短时过零率阈值设置 0.3，滑动窗口的长度设置 200（帧长度，25ms）

时，对语音信号“ot ket ken”（着火了）进行的初始词素端点检测。图中可以看出，含有三个词素的信号分割成 10 个语音段。我们对初始分割算法进行优化，得到以下图 4 所示的分割结果。

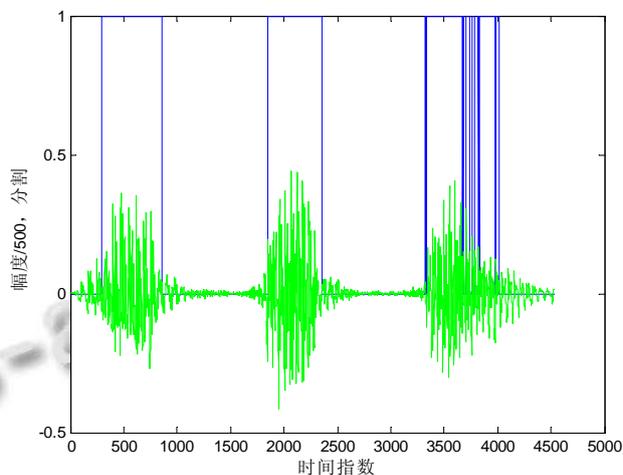


图 3 “ot ket ken”浊音/清音初始分割

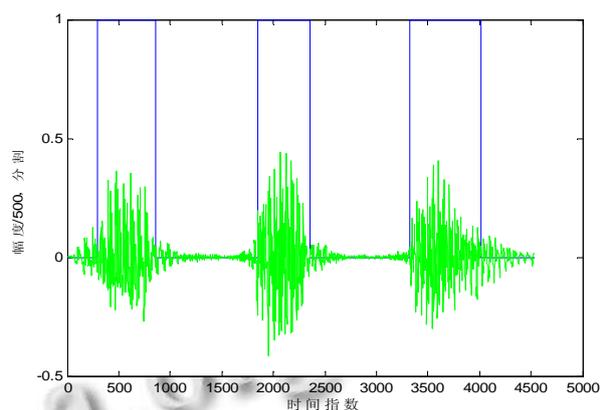


图 4 “ot ket ken”浊音/清音优化分割

图 4 是，我们根据维吾尔语词素的特征，对浊音/清音分割算法进行优化后得到的词素端点检测结果。图 4 表明，含有 3 个词素的语音信号正确分割成 3 个语音段。

7.3 特征提取

首先，把语音要进行数字化。数字化过程分两步：抽样和量化。一个信号通过度量它在特定时刻的振幅来进行抽样。抽样率就是每秒提取的样本数目。为了精确地测量声波，每周需要至少有两个样本：一个样本用于测量声波的正侧部分，一个样本用于测量声波的负侧部分。可能测量的最大频度的波就是那些频度等于抽样率一半的波。对于 8000Hz 的抽样率，要求每

秒测量 8000 个振幅。

我们用 LP 参数来提取每个帧的特征矢量序列。对敏感词“ot”（火）进行 LPC 特征提取（帧长 25ms，帧移 10ms），如图 5 所示。每个帧由 12 个倒谱系数和 12 个倒谱系数差组成。可以发现，倒谱系数是相互依赖的，大部分矢量的前 3 个特征数是一样的。倒谱系数差的相互依赖程度不是那么高。

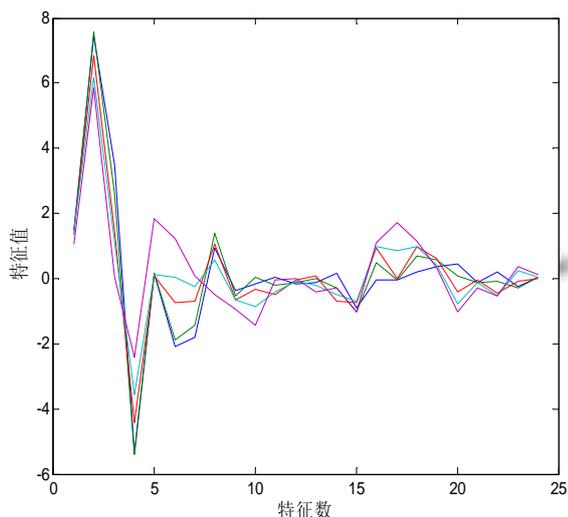


图 5 “ot”的特征

图 5 中可以看到，这里有 5 个特征矢量，表示 5 个重叠的帧。特征数 1~12 表示倒谱，特征数 13~24 表示倒谱差。

图 6 中可以看到，帧与帧之间的倒谱系数变化比较慢。图 7 显示，帧与帧之间的倒谱系数差变化比较快。

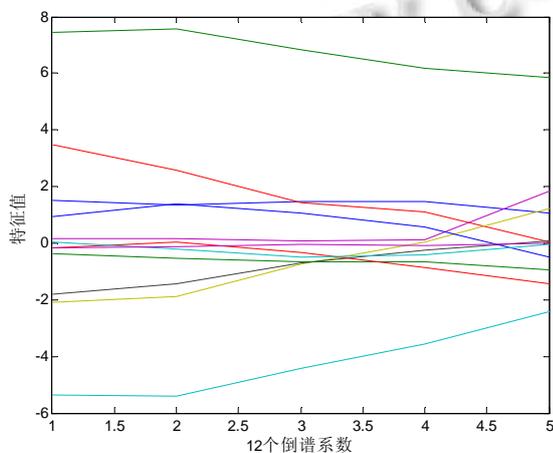


图 6 “ot”的倒谱系数

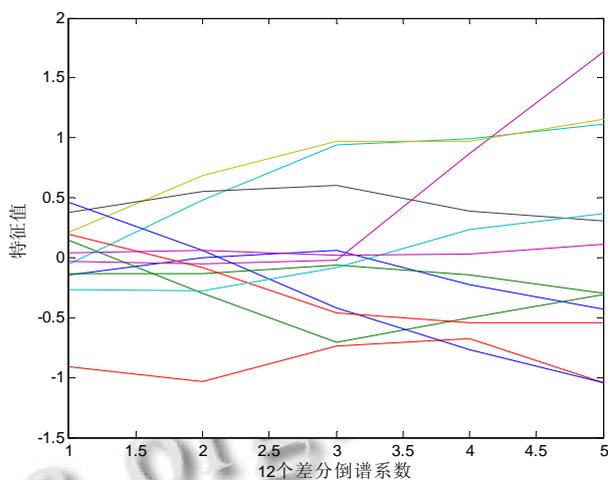


图 7 “ot”的分差倒谱系数

7.4 矢量量化

前面介绍了语音如何通过信号处理变换并转换成一系列特征矢量。每个特征矢量代表语音的一个时间片断。怎样把这些特征矢量转变成概率呢？特征矢量概率的一种计算方法是：首先进行聚类，把它们聚类为能够计数的离散符号，然后计算它们在某个训练集中出现的次数。只要对次数进行计数，就可以计算出给定聚类的概率。这种方法一般称为矢量量化^[5]。我们用 K-均值聚类方法来进行矢量量化。对每个敏感词，特征提取构成 24 维矢量的观察序列。然后对这个观察序列进行矢量量化生成离散的标量。

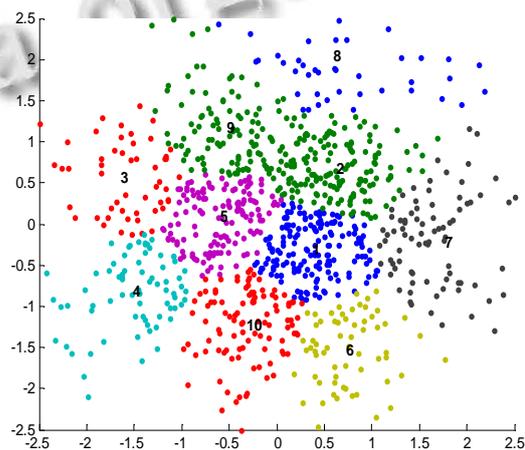


图 8 对随机产生的 1000 点的矢量量化

为了可视化的展现 K-均值函数矢量量化过程，我们用随机产生的 2 维矢量来进行 K-均值函数聚类，因

为用 2 维空间更客观的展现矢量量化过程。图 8 是对随机产生的 1000 点矢量量化聚类结果。聚类中心用数字来表示，属于同一个聚类的点用相同颜色来表示。

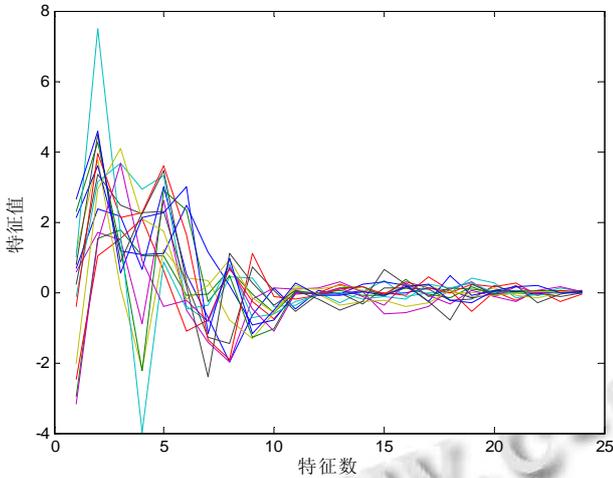


图 9 三个敏感词生成的码本

对每个敏感词，先进行基于帧的分析，获得观测矢量，然后串连所有敏感词的特征矢量并进行矢量量化生成 K 个特征矢量码本。我们对三个敏感词的训练集语料库生成一个码本， $K \leq 16$ ，如图 9 所示。

图 9 中显示，生成 15 个矢量原型（码本）。我们可以看到，矢量原型不像从单个单词提取的特征矢量那样互相依赖。覆盖所有训练集，特征矢量显示较多的变化并彼此之间保持一定的距离。衡量码本质量的方法是一个观测矢量在训练数据中与相应的符号的平均距离。我们把 K 分别设 2, 4, 8, 16, 32 等值，来分析失真与码本大小的对应关系，如图 10 所示。

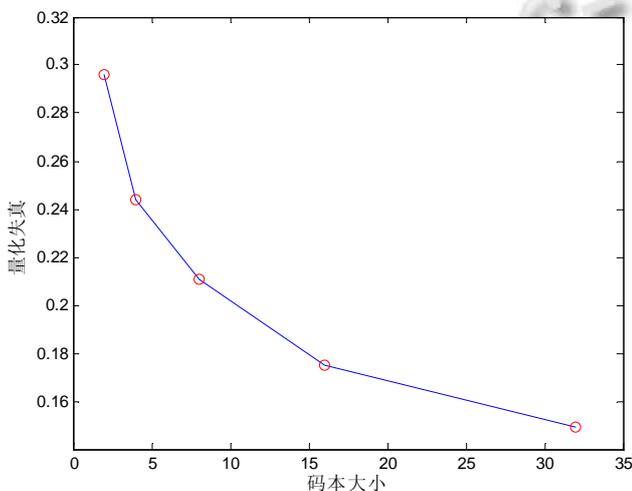


图 10 量化失真与码本大小对应关系

图 10 是，对含有三个敏感词，每个敏感词重复发音 2 次的训练集语料库进行量化失真与码本大小之间的关系图。图 10 中可以看到，码本大小越大失真度越低。因为较大的码本意味着更多的计算量，所以有一个激励，无危害的识别能力的前提下，使用尽可能小的码本。按照这两个情况，合理的码本大小由实验数据来确定。当矢量量化时，失真与码本对应曲线足够平的时候，就是最佳码本大小。

7.5 HMM 训练

对敏感词进行 HMM 训练的含义是用这个训练词汇估算状态转移矩阵 A ，观察概率矩阵 B 和状态概率矢量 π 。其目的是使某个 HMM 的概率最大化，使其产生它所代表的敏感词。这是迭代调用 FB 算法来实现的。还需要对同一个敏感词进行多次训练。用上面生成的码本进行基于特征提取、矢量量化和 FB 算法的迭代式、多层 HMM 训练。每个 HMM 模型有 5 个隐状态。图 11 是，建立敏感词“ot”的 HMM 模型时计算的对数似然和迭代数的关系图。

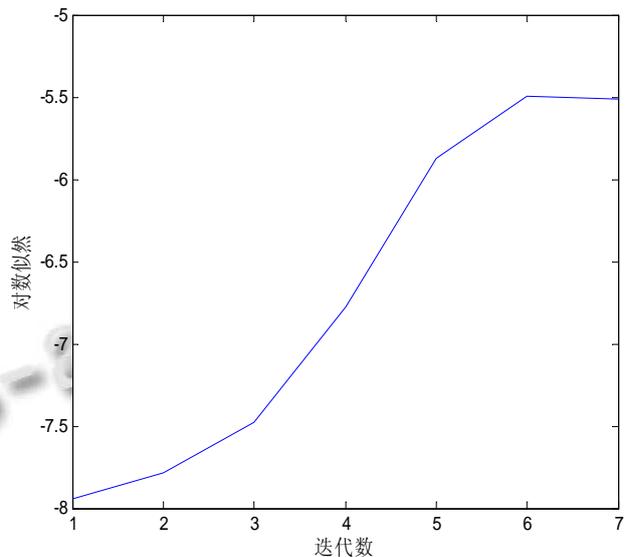


图 11 “ot”的对数似然和迭代数

图 11 中可以看到，本次训练要建立敏感词“ot”的 HMM 模型要用的迭代数为 7。通过实验可以知道，典型的训练长度应该在 20 次迭代左右。

我们也可以观察到矩阵状态转移概率分布矩阵 A 和符号的概率分布矩阵 B 。图 12 是，敏感词“ot”的两次发音序列 HMM 训练的 5×5 状态观察状态转移概率分布矩阵 A 。

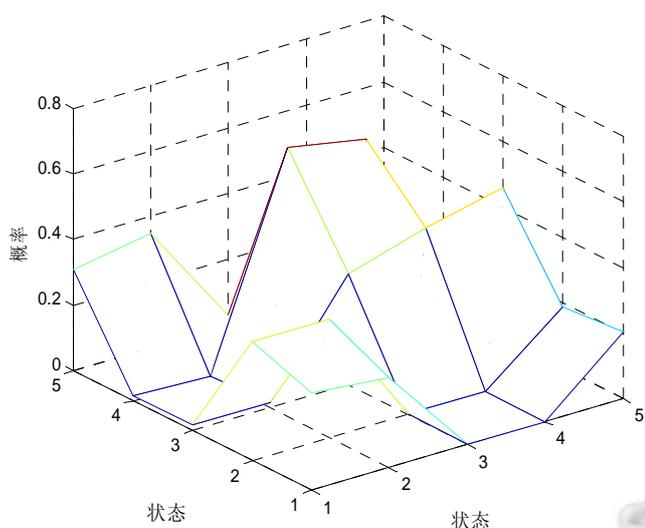


图 12 状态转移概率分布矩阵 A

图 12 中，可以观察到，矩阵 A 在主对角线的概率最大。这表明，如果进程在 t 时刻处某一个状态，那么图中概率最高的状态是在 $t+1$ 时刻的状态。故状态不会随时间迅速变化。

图 13 是，敏感词“ot”的两次发音序列 HMM 训练的 15×5 状态观察符号的概率分布矩阵 B。

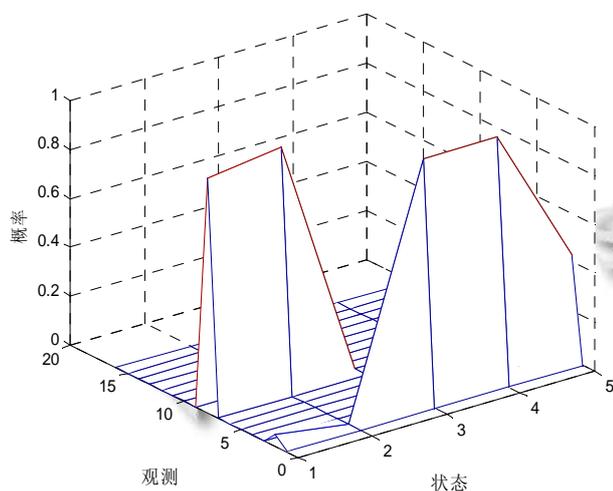


图 13 观察符号的概率分布矩阵 B

从图 13 中可以观察到，对每个状态，都有一对观测具有较高的概率。通常，一个状态有一个最高概率。比如，状态 3 和状态 4 在观测符号 2 的时候具有最高的概率。状态 1 在观测符号 7 和符号 9 的时候具有最高的概率。我们得出结论，隐马尔可夫模型不是完全

隐含的。

根据观察状态转移概率分布矩阵 A 与观察符号的概率分布矩阵 B 的结构和语音信号特征提取时得出的结论，如果当前状态的停留机会越大，并对每个状态观测到一个符号的可能性越大，则特征矢量随时间变化较慢。

7.6 HMM 语音识别

经过上面的训练，每个敏感词都生成一套模型参数。类似于训练集数据的矢量量化，将用于测试的敏感词的特征矢量送入第二步生成的矢量量化器。这样，用于测试的敏感词也是一个码矢编号序列，每个序列都经过所有训练敏感词的 HMM 模型参数计算概率值，这个过程采用 Viterbi 解码算法。这个算法使用的判据是寻找单个最佳状态序列(路径)，即使 $P(Q|O, ?)$ 最大，这等效于 $P(Q, O|?)$ 最大。这样概率最大值所对应的模型即为识别结果。最后用识别正确的敏感词数与所有测试集敏感词数做比值即为识别率。

给定一个观察(未知)序列和一定数量敏感词的 HMM 训练模型，我们能计算 HMM 产生的一序列对数似然。要识别一个给定信号，首先提取特征矢量，然后计算给定信号和给定 HMM 训练模型的对数似然，最后我们搜索出产生给定信号的概率最高的 HMM 训练模型。

我们对未知的维吾尔语广播新闻连续语音段“ot ket ken”进行识别。识别结果显示，输入连续语音信号“ot ket ken”中的从第 296 个采样点到第 858 个采样点之间的语音段与 HMM 训练模型中的敏感词“ot”的概率最大，即-0.0686。也就是说连续语音信号“ot ket ken”中，296~858 采样点出现敏感词“ot”。

```
logp =
-0.0686   -61.8361  -108.7968
-77.4896   -8.6842   -57.2943
-77.4896   -5.6495   -5.7365
ans =
'ot'      '296~858'
```

8 总结

本文以 Matlab 为开发平台，对来自于新疆维吾尔广播电台网站的维吾尔语新闻 60 分节目语音进行连续敏感词检索。首先，从新闻 60 分节目连续语音中手

(下转第 71 页)

- /dynamic spectrum access/cognitive radio wireless networks: a survey. *Computer Networks J*, 2006,50(13):2127–2159.
- 4 Clancy TC, Goergen N. Security in cognitive radio networks: threats and mitigation, 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2008), Singapore, May. 2008:5.
 - 5 Chen RL, Jung MP, Reed JH. *IEEE Journal on Selected Areas in Communications*, 2008,26(1):25–37.
 - 6 Chetan N, Mathur K, Subbalakshmi P. Security Issues in Cognitive Radio Networks. *Cognitive Networks*. 25 July 2007: 272–290.
 - 7 Bian KG, Park JM. MAC-Layer Misbehaviors in Multi-Hop Cognitive Radio Networks. 2006 US-Korea Conference on Science, Technology, and Entrepreneurship (UKC2006), Aug 2006.
 - 8 Brown TX, Sethi A. Potential Cognitive Radio Denial-of-Service Vulnerabilities and Protection Countermeasures: A Multi-dimensional Analysis and Assessment. *Cognitive Radio Oriented Wireless Networks and Communications, 2007. CrownCom 2007. 2nd International Conference on*. 1–3 Aug. 2007: 456–464.
 - 9 Sampath A, Dai H, Zheng HT, Zhao BY. Multi-channel Jamming Attacks using Cognitive Radios. *Computer Commu-*
 - nications and Networks, 2007. ICCCN 2007. Proc. of 16th International Conference on. 13–16 Aug. 2007: 352–357.
 - 10 Safdar GA, O'Neill M. Common control channel security framework for cognitive radio networks, 2009 IEEE 69th Vehicular Technology Conference, April 2009, Barcelona, Spain: 211–217.
 - 11 Haq A, Naveed A, Kanhere SS. Securing Channel Assignment in Multi-Radio Multi-Channel Wireless Mesh Networks, *Wireless Communications and Networking Conference 2007 WCNC*, Kowloon, March 2007: 3111–3116.
 - 12 Nishitan Y, Igarashi Y. Secure multi-party computation over networks. *IEICE Trans. on Information and Systems*, March 2000, E83–D(2):561–570.
 - 13 Asokan N, Ginzboorg P. Key agreement in ad hoc networks. *Computer Communications*, 2000,23(17):1627–1637.
 - 14 Yao AC. Protocols for secure computations, *Foundations of Computer Science*, 1982. SFCS'08. 23rd Annual Symposium on. Chicago USA, Nov 1982: 160–164.
 - 15 Rabin MO. How to exchange secrets with oblivious transfer, technical report TR–81. Aiken Computation Lab, Harvard University, 1981.

(上接第 35 页)

工截取同一个播音员、在不同的语音段发音的敏感词，共截取 3 个敏感词 (ot/火, yarlinish/伤, olush/死)，每个敏感词重复发音两次，建立训练语音语料库。然后，对同一个播音员的含有敏感词的另外一个新闻 60 分节目连续语音（不是截取训练敏感词的语音段）进行敏感词检索。

实验结果显示，一、帧长和帧移越大错误率越高，但识别速度越快、鲁棒性相当好。二、倒谱系数和分差倒谱系数越小识别的成功率越低，系统也越不稳定。三、特征矢量聚类数目越小识别率越低。

总之，帧长，帧移，倒谱系数，分差倒谱系数，码本大小，聚类数目等参数直接影响系统的识别率。所以识别率和识别速度之间取舍时，经实验来决定。

参考文献

- 1 哈力克·尼亚孜. 基础维吾尔语. 乌鲁木齐: 新疆大学出版社, 1997.
- 2 那斯尔江·吐尔逊, 吾守尔·斯拉木. 基于隐马尔可夫模型的维吾尔语连续语音识别系统. *计算机应用*, 2009, 29(7).
- 3 王炳锡, 屈丹, 彭焯, 等. 实用语音识别基础. 北京: 国防工业出版社, 2005.
- 4 屈丹, 王波, 李弼程, 等. VoIP 语音处理与识别. 北京: 国防工业出版社, 2010.
- 5 Jurafsky D, Martin JH. *Speech and Language Processing. Second Edition*, Beijing: Posts & Telecom Press, 2010.
- 6 Hahn B, Valentine D. *Essential Matlab for Engineers and Scientists. Fourth Edition*, Canada: Academic Press, 2010.
- 7 张德丰, 等. *Matlab 概率与数理统计分析*. 北京: 机械工业出版社, 2010.