高技术通讯 2009 年 第 19 卷 第 2 期:119 ~ 124

doi: 10.3772/j.issn.1002-0470.2009.02.002

基于后验概率的汉语语音检索方法研究①

郑铁然② 韩纪庆

(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

摘 要 针对经典的向量空间检索模型直接用于基于音节 lattice 形式的汉语语音检索存在无法有效区分 lattice 中包含的正确音节识别候选和错误的识别候选以及不能充分利用 lattice 中所蕴含的各层级信息的不足,提出了一种基于语音文档邻接音节后验概率矩阵的检索方法。该方法以该矩阵作为文档索引,并计算查询请求被包含在语音文档中的后验概率,并以此来度量查询请求和语音文档间的相关度。后验概率作为可靠的置信测度能够有效区分正确和错误的音节候选,在 lattice 中后验概率的计算能够充分地利用语音识别结果中的多层级的信息。语音检索实验表明,与基于向量空间模型的检索方法相比,该方法的检索性能有显著提高,是一种适用于汉语语音检索的有效方法。

关键词 汉语语音检索, 音节 lattice, 后验概率, 检索模型, 邻接矩阵

0 引 言

随着计算机技术和多媒体技术的发展,被人们记录并保存在计算机中的语音数据越来越多。为了更高效地管理和利用这些语音资源,必须实现基于语义内容的快速语音检索。本文进行了基于音节网格(lattice)的汉语语音检索研究,提出了一种基于语音文档邻接音节后验概率矩阵的检索方法,该方法能够充分利用语音识别结果中各层次的信息,是一种非常有效的且适用于汉语语音检索任务要求的检索方法。

1 基本知识和研究背景

语音检索任务是指根据用户输入的查询请求(query)在语音资源中搜索和返回与之相关联的语音段或语音文件的处理过程。为了实现快速检索,这一过程一般要分成"离线"和"在线"两个阶段来完成。在"离线"阶段,通过自动语音识别(automatic speech recognition, ASR)技术识别语音数据,将其转化为对应的文本表示,并基于该文本表示建立索引。而在"在线"阶段则仅根据索引内容判定文档与查询请求之间的相关程度。

根据索引策略的不同,可以将已有的汉语语音检索研究工作分成三大类:第一类是基于大词表连续语音识别(large vocabulary continuous speech recognition, LVCSR)系统的识别结果来建立索引[1];第二类是基于子词识别器所产生的子词文本来建立索引[2-4];第三类是基于子词识别器的 lattice 识别结果来建立索引[5-7]。Lattice 是语音识别器最常采用的多候选输出形式,它用有向图的形式实现了对多个候选结果的精简表示。

LVCSR 的识别结果总受限于词表内容,因而第一类检索系统不能很好地支持查询请求中的词表外(out-of-vocabulary, OOV)词,这使得研究者不得不转向基于子词的检索方法。在第二类语音检索研究中,虽然广泛地尝试了各种层次的子词基元,如音素[2,3]、广义音素类[2]、自动选择的"元音-辅音-元音"组[4]、音节[2]、音素 M-grams[2]等,但研究表明,子词识别结果较高的识别错误率使得检索性能受到严重制约。相比较而言,基于子词 lattice 的检索方法更加可靠也有更好的发展空间,一方面采用多候选形式有效地补偿了识别错误对检索性能的影响,另一方面 lattice 中所承载的各种信息是实现高效检索的有力依据。对汉语而言,音节因其数量有限且独立性强成为最常采用的子词基元形式[8]。目前,基于音节 lattice 的检索方法是汉语语音检索研究的主

① 国家自然科学基金(60575030)和 863 计划(2006AA01Z197)资助项目。
② 男,1972 年生,博士生;研究方向:语音检索,语音识别等;联系人,E-mail: zhengtieran@hit.edu.cn
(收稿日期:2007-11-27)

流方法。

在基于音节 lattice 的汉语语音检索研究中，检索模型一般采用向量空间模型(vector space model，VSM)。典型的如文献[8]所示，采用音节和音节对为特征索引项，以声学得分为权值来构建文档向量，并用余弦相似度来度量查询请求与文档间的相关度。近几年来，研究者在向量空间模型的基础上尝试引入查询扩展(query expansion)[9]和潜层语义分析(latent semantic analysis，LSA)[10]等技术以实现概念层次的检索。

本文认为，在基于音节 lattice 的语音检索中直接应用向量空间检索模型有很多不足：(1) lattice 中不仅包含正确的音节识别候选，也包含许多错误的识别候选，这两部分内容在向量空间模型中无法得到有效区分；(2)向量空间模型所构建的文档向量只能以 lattice 为一个整体来反映其组成内容，对 lattice 中所蕴含的其它各种层级的信息如声学概率、语音学概率及多候选竞争关系等，则未能加以充分利用。这种停留在表面层次的利用割裂了语音识别和语音检索这两部分工作间的有效衔接，严重制约了检索性能。针对这些不足，本文提出了一种基于语音文档的邻接音节后验概率矩阵的检索方法，该方法直接计算查询请求被包含在语音文档中的后验概率，并以此来度量查询请求和语音文档间的相关程度。一方面后验概率这种可靠的置信测度手段，能够有效区分正识和错识内容[11]，另一方面在 lattice 中后验概率的计算充分地利用了各层级的信息，从而实现了更加有效的检索。

2 语音文档和 lattice 的表示

语音文档 d 可以看作由语句 O 组成的序列，$d = O_1\ O_2 \cdots O_N$。设语句 O_n 的时长为 T_n。离线索引时，用音节识别器识别每个语句 O_n，且为其生成一个 lattice 形式的识别结果 G_n。

G_n 可以表示为一个有向图 (V_n, E_n)，其中 V_n 是顶点集，集中每个顶点 v 都是一个音节候选，它有 3 个属性：起始时间 $s_n(v)$、结束时间和声学得分。E_n 是边集，其中的每一个边 $e = (u, v)$ 描述了顶点 u 和顶点 v 间的邻接关系。每边拥有属性 $q_n(u, v) = P(v \mid u)$，表示在音节 u 之后识别到音节 v 的概率，即语言模型得分。若有向图 G_n 中的一条路径 $w = v_1\ v_2 \cdots v_M$ 存在性质 $s_n(v_1) = 0, t_n(v_M) = T_n$，则称 w 为 G_n 的一条全局路径，它所对应的音节串就是语句 O_n 的一个识别结果候选。由于在 lattice 中总是存在多条全局路径，因此就构成了识别结果的多候选表示。记所有的全局路径的集合为 W_n，包含邻接顶点 v_iv_j 的全局路径的集合为 $W_n(v_iv_j)$，包含音节对 s_is_j 的全局路径的集合为 $W_n(s_is_j)$。

语音文档 d 的识别结果候选可以认为是由语句 O 的识别结果候选组合而成。序列 $c = w_1\ w_2 \cdots w_N$，有 $w_n \in W_n$，则称 c 是语音文档 d 的一个识别结果候选，对应一个可能的文档内容。这样的文档内容当然也有多个候选，记其全集为 C，同样记包含有音节对 s_is_j 的文档内容的集合为 $C(s_is_j)$。

3 lattice 的邻接音节后验概率矩阵

有向图可以表示成邻接矩阵的形式。我们借鉴这一思路，也用邻接矩阵的形式表示 lattice，并采用后验概率作为矩阵中各元素的权值。首先构建基于邻接顶点的后验概率矩阵，然后将其变化成邻接音节的后验概率矩阵形式。这样的后验概率矩阵将不仅反映顶点或音节间的邻接关系，而且对邻接关系进行了有效的置信测度。第一步需要计算在 O_n 中包含邻接顶点对 v_iv_j 的后验概率 $P(v_iv_j \mid O_n)$。

3.1 计算 $P(v_iv_j \mid O_n)$

有：

$$\begin{aligned} P(v_iv_j \mid O_n) &= P(W_n(v_iv_j) \mid O_n) \\ &= \frac{P(W_n(v_iv)_j, O_n)}{P(O_n)} \\ &= \frac{\sum\limits_{w \in W_n(v_iv_j)} P(w, O_n)}{\sum\limits_{w \in W_n} P(w, O_n)} \\ &= \frac{\sum\limits_{w \in W_n(v_iv_j)} P(O_n \mid w)P(w)}{\sum\limits_{w \in W_n} P(O_n \mid w)P(w)} \end{aligned} \tag{1}$$

若设全局路径 $w = v_1\ v_2 \cdots v_M$，则路径的声学模型概率 $P(O_n \mid w)$ 和语言模型概率 $P(w)$ 可以如下计算：

$$P(O_n \mid w) = \sum_{m=1}^{M} b_n(v_m) \tag{2}$$

$$P(w) = \sum_{m=2}^{M} q_n(v_{m-1}, v_m) \tag{3}$$

式中声学得分 $b_n(v_m)$ 和语言模型得分 $q_n(v_{m-1}, v_m)$ 是承载在 lattice 中的已知信息，所以显然式(1)是可以计算的。但直接计算需要遍历所有的全局路径，

计算量非常大,因而需要快速算法。

借鉴文献[5]中所述的后验概率计算方法,设计求取 $P(v_iv_j \mid O_n)$ 的动态规划算法。若称起始时间为 0 的顶点为 lattice 的首顶点,称结束时间为 T_n 的顶点为 lattice 的尾顶点。则首先定义:从首顶点出发到顶点 v 终止的所有局部路径的集合为 $W_n^-(v)$;从顶点 v 出发到尾顶点终止的所有局部路径的集合为 $W_n^+(v)$;语句 O_n 中从 t_1 时刻到 t_2 时刻的语音段为 $O_n(t_1,t_2)$。然后定义顶点 v 前向变量和后向变量 $\beta_n(v)$ 如下:

$$\begin{aligned}\alpha_n(v) &= P(W_n^-(v), O_n(0,t_n(v))) \\ &= \sum_{w\in W_n^-(v)} P(w, O(0,t_n(v)))\end{aligned} \tag{4}$$

$$\begin{aligned}\beta_n(v) &= P(W_n^+(v), O_n(s_n(v),T_n)) \\ &= \sum_{w\in W_n^+(v)} P(w, O(s_n(v),T_n))\end{aligned} \tag{5}$$

$\alpha_n(v)$ 和 $\beta_n(v)$ 具有如下性质:

(1) 初值。若 v 是首顶点,即有 $s_n(v)=0$,则有

$$\alpha_n(v) = b_n(v) \tag{6}$$

若 v 是尾顶点,即有 $t_n(v)=T_n$,则有

$$\beta_n(v) = b_n(v) \tag{7}$$

(2) 递推关系。对所有的 $v,u\in V_n$ 有

$$\alpha_n(v) = \sum_{(u,v)\in E_n} \alpha_n(u)q_n(u,v)b_n(v) \tag{8}$$

$$\beta_n(u) = \sum_{(u,v)\in E_n} b_n(u)q_n(u,v)\beta_n(v) \tag{9}$$

(3) 可以用于计算 $P(O_n)$ 和 $P(v_iv_j \mid O_n)$:

$$\begin{aligned}P(O_n) &= \sum_{w\in W_n} P(w, O_n(0,T_n)) \\ &= \sum_{v\in V_n \cap t_n(v)=T_n} \alpha_n(v)\end{aligned} \tag{10}$$

将式(2)和式(3)代入到式(1)分式的分子部分,并在其累计求和式中提取公共部分 $q_n(v_i,v_j)$,则式(1)可以变换为如下形式:

$$P(v_iv_j \mid O_n) = \frac{\alpha_n(v_i)q_n(v_i,v_j)\beta_n(v_j)}{P(O_n)} \tag{11}$$

利用这 3 个性质,首先计算首顶点的前向变量值和尾顶点的后向变量值,然后利用式(8)和式(9)的递推关系,分别向前和向后逐步递推,直到求出所有顶点的前向变量值和后向变量值,最后根据式(10)用所用尾顶点的前向变量值求出 $P(O_n)$,根据顶点 v_i 的前向变量值和顶点 v_j 的后向变量值以及 $P(O_n)$,利用式(11)求出 $P(v_iv_j \mid O_n)$。

3.2 lattice 的邻接顶点后验概率矩阵

在 lattice 中求出所有邻接顶点的后验概率,然后构建 G_n 的邻接顶点后验概率矩阵 $\boldsymbol{H_n} = \{h_{ij}^n\}_{L_n\times L_n}$,其中 L_n 为 G_n 的顶点总数量,h_{ij}^n 为顶点 v_i 和顶点 v_j 相邻接的后验概率,即有:

$$\begin{cases} h_{ij}^n = P(v_iv_j \mid O_n) & (v_i,v_j)\in E_n \\ h_{ij}^n = 0 & (v_i,v_j)\notin E_n \end{cases} \tag{12}$$

3.3 lattice 的邻接音节后验概率矩阵

从检索的角度,我们更关心的是音节间的邻接关系以及其置信测度,因而需要将邻接顶点后验概率矩阵变化为邻接音节后验概率矩阵。设全部音节的集合为 S,音节总数量为 K。首先构建 G_n 的顶点到音节的映射关系矩阵 $\boldsymbol{A_n} = \{a_{ij}^n\}_{L_n\times K}$。在 G_n 中,若顶点 v_i 所表示的识别候选是音节 s_j,则 $a_{ij}^n = 1$,否则 $a_{ij}^n = 0$。然后按下式计算 G_n 的邻接音节后验概率矩阵:

$$\boldsymbol{B_n} = \{b_{ij}^n\}_{K\times K}$$

$$\boldsymbol{B_n} = \boldsymbol{A_n'}\boldsymbol{H_n}\boldsymbol{A_n} \tag{13}$$

式中,$\boldsymbol{A_n'}$ 为 $\boldsymbol{A_n}$ 的转置矩阵。从式(13)推导可知:

$$b_{ij}^n = \sum_{k_1=1}^{L_n}\sum_{k_2=1}^{L_n}[a_{k_1i}^n P(v_{k_1}v_{k_2} \mid O_n) a_{k_2j}^n] \tag{14}$$

即 b_{ij}^n 表示为符合音节对 s_is_j 的所有邻接顶点的后验概率的累加值,它实际上等于集合 $W_n(s_is_j)$ 在 O_n 中的后验概率。即有 $b_{ij}^n = P(s_is_j \mid O_n)$。因此 lattice 的邻接音节后验概率矩阵 $\boldsymbol{B_n}$ 就反映了 O_n 识别结果中各音节的邻接关系及其后验概率。

4 语音文档的邻接音节后验概率矩阵

在求出了每个语句的邻接音节后验概率矩阵后,就可以计算整个语音文档 d 的邻接音节后验概率矩阵 $\boldsymbol{R} = \{r_{ij}\}_{K\times K}$。令矩阵元素 $r_{ij} = P(s_is_j \mid d)$,表示为在文档 d 中包含音节对 s_is_j 的后验概率。$P(s_is_j \mid d)$ 可以如下计算:

$$\begin{aligned}P(s_is_j \mid d) &= P(C(s_is_j) \mid d)) \\ &= \frac{\sum_{c\in C(s_is_j)} P(d \mid c)P(c)}{P(d)}\end{aligned} \tag{15}$$

设 $c = w_1\,w_2\cdots w_N, w_n\in W_n$。此时式(15)中的求和范围 $c\in C(s_is_j)$ 也可以写为如下逻辑表达式:$\bigcup_{n=1}^{N} w_n \in W_n(s_is_j)$。且可以认为各语句是相互独立的,其识别结果也是相互独立的,则有

$$P(c) = P(w_1\,w_2\cdots w_N) = \prod_{n=1}^{N} P(w_n) \tag{16}$$

$$P(d) = P(O_1\ O_2 \cdots O_N) = \prod_{n=1}^{N} P(O_n) \quad (17)$$

$$P(d \mid c) = P(O_1\ O_2 \cdots O_N \mid w_1\ w_2 \cdots w_N) = \prod_{n=1}^{N} P(O_n \mid w_n) \quad (18)$$

基于上述分析,式(15)可以变换成:

$$P(s_i s_j \mid d) = \sum_{\bigcup_{n=1}^{N} w_n \in W_n(s_i s_j)} \prod_{n=1}^{N} P(w_n \mid O_n) \quad (19)$$

进一步对式(19)进行变换,可以得到如下计算公式:

$$P(s_i s_j \mid d) = 1 - \prod_{n=1}^{N} [1 - P(s_i s_j \mid O_n)] \quad (20)$$

即 $P(s_i s_j \mid d)$ 可以通过 $P(s_i s_j \mid O_n)$ 计算得到。由于各语句的邻接音节后验概率矩阵已经求出,$P(s_i s_j \mid O_n)$ 为已知值,因而根据式(20)可以非常快速地计算 $P(s_i s_j \mid d)$,并最终构建语音文档 d 的邻接音节后验概率矩阵 $\boldsymbol{R}$。这个后验概率矩阵表示了语音文档中各音节的邻接关系和其置信测度。

5 基于邻接音节后验概率矩阵的语音检索方法

在语音检索中,查询请求往往采用文本形式提交,并由若干个查询词构成。文档集中那些包含全部查询词,且各词有较高出现频率的语音文档一般会被认为与查询请求间有较高的相关度。本文的检索方法就是针对这种类型的语音检索任务提出的。

在"离线索引"阶段,计算语音文档 d 的邻接音节后验概率矩阵 $\boldsymbol{R}$,并存储它作为文档 d 的索引。在"在线检索"阶段,需要通过 $\boldsymbol{R}$ 来计算语音文档 d 和查询请求 q 间的相关度 $SIM(d,q)$。首先将查询请求 q 分解为若干查询词,并找到组成查询词的所有邻接音节对,并设这些音节对的集合为 Q。集合 Q 可以看作是查询请求 q 的另一种表现形式,而根据索引矩阵 $\boldsymbol{R}$,可以计算集合 Q 被包含在语音文档 d 中的后验概率值 $P(Q \mid d)$。

$$P(Q \mid d) = \prod_{s_i s_j \in Q} P(s_i s_j \mid d) \quad (21)$$

计算时假定了集合中各音节对是相互独立的。式中的 $P(s_i s_j \mid d)$ 可以通过查找语音文档 d 的邻接音节后验概率矩阵 $\boldsymbol{R}$ 的对应元素来得到。我们认为这一后验概率值合理地度量了查询请求和语音文档间的相关度,即令

$$SIM(d,q) = P(Q \mid d) \quad (22)$$

在整个文档集上计算每个语音文档与查询请求间的相关度,并基于这一相关度对语音文档进行排序,从而得到检索结果。在实际设计检索算法时,为了避免零概率带来零相关度这种现象出现,建立索引时可以将矩阵 $\boldsymbol{R}$ 中所有零元素设定为一个很小的非零固定值,本文设定为 0.0001。

6 实验及结果分析

6.1 测试语料和识别器

测试语料采集自 CCTV 的"新闻联播"、"新闻会客室"、"午间新闻"、"焦点访谈"等节目,时长 40 小时,手工进行了音节标注,并注明了文档边界和句子边界。语料中共包含 632 篇语音文档,音节总数为 529923 个。文档长短分布如表 1,其中最短文档 42 个音节,最长文档 41637 个音节。测试用的查询请求都采用单个词的形式,由 30 个二字词和 20 个三字词组成了测试 Query 集,手工注明每个词与测试语音文档间的相关性。

表 1 测试语料中语音文档的长短分布

长度类型	文档数
短(音节数 < 200)	81
中(200 ≤ 音节数 ≤ 1000)	494
长(音节数 > 2000)	57

构造两个音节识别器来分别生成基于有调音节的识别结果和基于无调音节的识别结果,分别涵盖 1677 个有调音节和 418 个无调音节。训练语料来自 863 语料库,声学模型采用基于声韵结构的 Triphone 模型,语言模型采用基于音节的 Bi-gram 模型。采用最大后验概率(maximum a posteriori, MAP)算法在 1 小时测试语料上进行了自适应。在整个测试集合上,有调音节识别器的 One-Best 识别结果的音节识别率为 62.3%,无调音节识别器的 One-Best 识别结果的音节识别率为 71.6%。

6.2 不同检索方法的检索性能比较

在测试语料上,采用音节识别器为每个语句都生成一个 lattice 识别结果,识别时设定 Beam 剪枝宽度为 350。在这些 lattice 上,分别采用基于向量空间模型的检索方法和本文所述的基于语音文档邻接音节后验概率矩阵的检索方法进行语音检索实验。其中,基于向量空间模型的检索方法,以全部有调音节和音节对为特征索引项,并分别采用词频-逆向文件频率(term frequency-inverse document frequency, TF-

IDF)词频[12]和声学得分累计值[8]两种不同的索引项权值计算方法来构建文档向量。检索时,计算文档向量与查询请求向量间的余弦相似度作为相关性度量。而本文的方法也是在有调音节识别结果上实现的。实验中,逐一检索全部 50 个查询请求,然后对检索性能进行综合评价。图 1 所示的查准率(precision)/查全率(recall)曲线比较了各检索方法的性能。

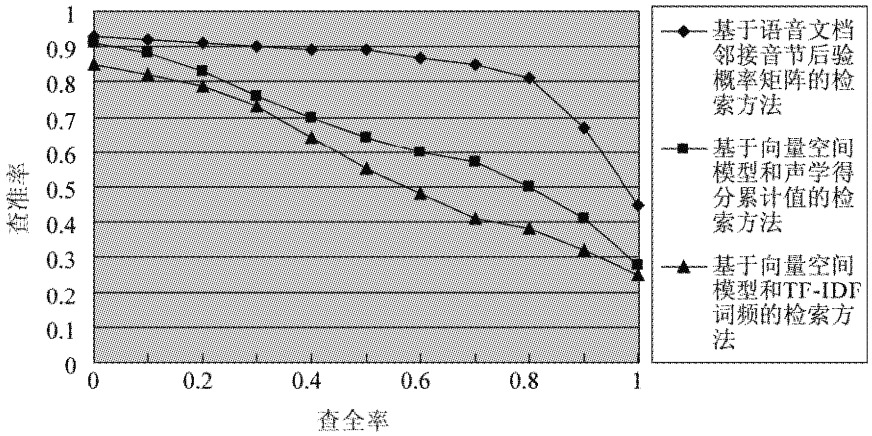

图 1　不同检索方法的查准率/查全率曲线

从图 1 可知,两种基于向量空间模型的检索方法的性能曲线都呈现了这样一个趋势:即随着查全率增加,查准率急剧下降。这表明不论采用何种权值计算方法,其构造的文档向量都存在着一定的偏差,从而影响了检索性能。而本文提出的基于语音文档邻接音节后验概率矩阵的检索方法的性能曲线变化则平缓得多,且在相同的查全率条件下,查准率值要远远好于其它两条曲线。这证明了我们提出的基于语音文档邻接音节后验概率矩阵的检索方法是有效的。与基于向量空间模型的检索方法相比较,该方法因为能够有效地区分 lattice 中的正确候选和错误候选,且能够更充分地利用语音识别结果中各层级的信息,从而使检索性能有了非常显著的提高。

6.3　采用不同音节基元形式时的检索性能比较

从数学模型上看,本文的方法不仅适用于有调音节,也适用于无调音节或其它子词形式。在基于有调音节的识别结果上,以有调音节为索引基元建立邻接音节后验概率矩阵,在基于无调音节的识别结果上,以无调音节为索引基元建立邻接音节后验概率矩阵,并分别进行检索实验。图 2 所示的查准率/查全率曲线比较了它们的检索性能。

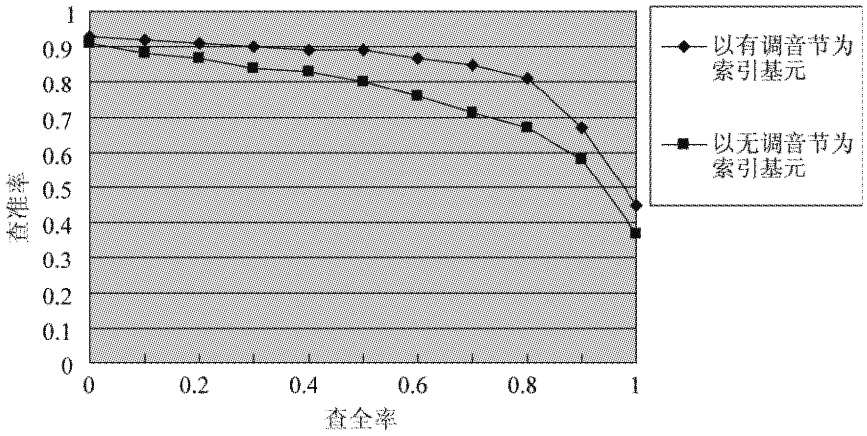

图 2　采用不同音节基元形式时的检索性能比较

从图 2 可知,采用无调音节为索引基元时的检索性能要比采用有调音节时差。分析原因如下:(1)虽然无调音节识别结果的 one-best 识别率相对较高,但本文的方法由于能够有效区分 lattice 中的正

确与错误音节候选,因而实际上是基于 lattice 中最优识别候选来得到检索结果的,而在 beam 剪枝宽度为 350 的条件下,两种 lattice 的最优候选的识别率差距很小;(2) 采用无调音节时,由于无法区分同音不同调的语音内容,因而误检较多影响了检索性能。

7 结 论

本文在基于音节 lattice 的汉语语音检索研究中,提出了一种基于语音文档邻接音节后验概率矩阵的检索方法。语音检索实验表明,其性能要远远优于基于向量空间模型的检索方法。该方法能够充分地利用语音识别结果中各层级的信息,是一种非常有效且适用于汉语语音检索任务要求的检索方法。

参考文献

[1] Abberley D, Renals S, Cook G. Retrieval of broadcast news documents with the THISL system. In: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing, Seattle, 1998. 3781-3784

[2] Ng K, Zue V W. Subword-based approaches for spoken document retrieval. *Speech Communication*, 2000, 32:157-186

[3] Beth L, Pedro M, Om D. Word and subword indexing approaches for reducing the effects of OOV queries on spoken audio. In: Proceedings of the 2002 Human Language Technology Conference, San Diego,2002

[4] Wechsler M, Schauble P. Speech retrieval based on automatic indexing. In: Proceedings of the Final Workshop on Multimedia Information Retrieval, Glasgow, 1995

[5] Seide F. Vocabulary independent search in spontaneous speech. In: Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, 2004. I253-I256

[6] Saraclar M, Sproat R. Lattice-based search for spoken utterance retrieval. In: Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics 2004, Boston, Massachusetts, USA, 2004. 129-136

[7] Hori T, Hetherington I L, Hazen T J, et al. Open-vocabulary spoken utterance retrieval using confusion networks. In: Proceedings of the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing, Honolulu, HI, USA, 2007. 4: 73-76

[8] Bai B R, Chen B L, Wang H M. Syllable based Chinese text/spoken document retrieval. *International Journal of Pattern Recognition and Artificial Intelligence*, 2000, 14(5): 603-616

[9] Logan B, Thong V, Moreno P J. Approaches to reduce the effects of OOV queries on indexed spoken audio. IEEE transactions on multimedia, 2005, 7: 899-906

[10] Chen B. Exploring the use of latent topical information for statistical Chinese spoken document retrieval. *Pattern recognition Letters*, 2006, 27(1): 9-18

[11] Soong F K, Wai-Kit Lo, Nakamura S. Generalized word posterior probability (GWPP) for measuring reliability of recognized words. In: Proceedings of the Special Workshop in Maui 2004, Maui, Hawaii, 2004.127-128

[12] Baeza Y R, Ribeiro N B. Modern Information Retrieval. Massachusetts: Addison Wesley Longman Publishing Company, 1999. 27-30

Study on Chinese speech retrieval based on posterior probability

Zheng Tieran, Han Jiqing

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

Abstract

In view of the fact that the syllable lattice based Chinese speech retrieval with the vector space retrieval model lacks the ability to effectively distinguish the incorrect syllable recognition from the correct syllable recognition in the lattice and can not fully utilize the multilevel information in recongnition results, the paper proposes a retrieval method based on the neighbor syllable posterior probability matrix. The method takes the matrix as the speech indexes and calculates the posterior probability of including a query in speech, and based on this, acquires the correlation between the query and the speech. The results of the speech retrieval experiment show the method's excellent performance in Chinese speech retrieval compared with the method base on the vector space model.

Key words: Chinese speech retrieval, syllable lattice, posterior probability, retrieval model, neighbor matrix