

# 语音关键词识别中基于 MLP 帧级子词 后验概率的置信度方法

李文昕<sup>1</sup> 屈 丹<sup>1</sup> 李弼程<sup>1</sup> 刘 崧<sup>2</sup>

(1. 解放军信息工程大学 信息工程学院, 河南 郑州 450002; 2. 恒生数字有限公司, 浙江 杭州 310012)

**摘 要:** 针对关键词检测系统中 HMM 模型框架下置信度计算存在的不足, 本文提出了基于 MLP 帧级子词后验概率的置信度方法。与 HMM 模型框架下利用声学模型得分与语言模型得分进行置信度计算不同的是, 该方法在 MLP 模型框架下直接将其输出的每帧语音类别的后验概率用于关键词置信度的计算, 克服了 HMM 建模时假设每帧语音的声学特征相互独立以及对状态建模时采用有限混元的高斯分布的不足。关键词检出和置信度确认使用两套不同的模型结构, 是两个完全独立的过程, 便于融合其他的置信度特征。实验结果表明, 本文提出的方法优于 HMM 框架下主流的置信度计算方法, 且与其具有较好的互补性。因此本文将两种不同框架下不同的置信度方法进行融合, 系统的等错误率(EER)相对提高了11.5%。

**关键词:** 关键词检出; 置信度计算; 多层感知器; 后验概率

**中图分类号:** TP391.42    **文献标识码:** A    **文章编号:** 1003-0530(2012)07-1051-06

## Confidence Measure Based On Frame-level Sub-word Posterior Probability of MLP for Speech Keyword Recognition

LI Wen-xin<sup>1</sup> QU Dan<sup>1</sup> LI Bi-cheng<sup>1</sup> LIU Song<sup>2</sup>

(1. Institute of Information Engineering, Information Engineering University of PLA, Zhengzhou 450002, China;  
2. Handsome Digital Equipment Technology Company, Hangzhou 310012, China)

**Abstract:** As the confidence measures in the scheme of Hidden Markov Model (HMM) in keyword spotting system have some shortcomings, a confidence measure based on frame-level sub-word posterior probability of Multi-layer Perception (MLP) is presented in this paper. Conventionally, the confidence is calculated from the acoustic and language model scores computed by the recogniser of HMM model, which makes some incorrect assumptions, such as the frame-wise and possibly component-wise independence of acoustic features, and a finite number of Gaussian mixtures. The proposed confidence measure is directly calculated from the frame-level sub-word posterior probabilities produced by a MLP network. The confidence estimation is completely separated from the keyword spotting and they use two different models. With this separation, decision making can be addressed with more reliable confidence and multiple confidence features can be integrated to improve the decision quality. The experimental results show that the proposed approach in this paper is better than the mainstream confidence measures in the framework of HMM model and they have good complement, when combining with the mainstream confidence measures in the scheme of HMM model, the Equal Error Rate (EER) of keyword spotting system achieves 11.5% relative improvement.

**Key words:** Keyword spotting; Confidence Measure; Multi-layer Perception; Posterior probability

## 1 引言

语音关键词识别<sup>[1]</sup>是从连续的、无限制的自然语音流中识别出一组给定的词即关键词的过程,被认为是处理自然语言、实现人机语音交互最有效的解决方案之一,并成为语音识别领域的一个研究热点,在实时的对话系统、命令检测以及语音文档内容检索等领域都发挥着至关重要的作用。目前的关键词识别系统包括两种方法,一种是基于垃圾模型的关键词检测系统<sup>[2]</sup>(Keyword Spotting 系统),主要面向小词表、限定领域的对话系统和语音命令检测任务;另一种是基于 Lattice 的关键词检测系统<sup>[3][4]</sup>(Spoken Term Detection 系统),主要应用于与关键词表、任务无关的语音监听、音频信息检索等领域,以应对互联网上大量出现的音频信息。

关键词检测系统包括关键词检出和确认两大模块,关键词检出部分尽可能多的检测出候选关键词,降低系统的漏报,再由关键词确认部分对检出的关键词进行置信度打分,确认候选关键词的可信程度,降低系统的误报。置信度的区分能力在关键词检测系统中起着至关重要的作用。现有的置信度计算方法<sup>[5]</sup>主要集中在四个方面:一是基于特征的方法<sup>[6]</sup>,主要是寻找好的置信度特征。二是基于似然比检验的方法<sup>[7][8]</sup>,通常是计算声学模型的似然比,主要用在基于垃圾模型的关键词检测系统框架下。三是基于后验概率的方法<sup>[9][10]</sup>,主要用在基于 Lattice 的关键词检测系统框架下。四是基于区分性置信度的方法<sup>[11]</sup>,主要是选取合适的区分性模型直接计算分类的后验概率。目前,国内外各大研究机构广泛使用的是基于似然比检验和基于 Lattice 后验概率的置信度计算方法,这两种方法都是在 HMM 框架下利用声学模型得分、语言模型得分等识别结果本身的信息进行置信度计算,置信度的得分对识别结果的可靠程度不能够提供一个独立的评估。另外,HMM 模型假设每帧语音的声学特征是相互独立的,而声学置信度是一个局部得分,高度依赖语音的当前帧和相邻帧;HMM 对状态建模时采用的是有限混元的高斯分布,并不是语音的真实分布。在置信度计算过程中,基于 Lattice 后验概率的置信度计算方法需要利用整个 Lattice 上的得分信息,计算代价较大;基于似然比检验的方法需要计算反词模型的得分,反词模型的构建比较繁琐。而多层感知器(Multi-layer Perception, MLP)可以很好的解决 HMM 框架下置信度计

算所面临的这些问题。

MLP 是人工神经网络(Artificial Neural Network, ANN)的一种,用于计算分类类别的后验概率,在连续语音识别中应用广泛,主要与 HMM 结合使用,将 MLP 输出各音素类别的后验概率作为 HMM 状态的输出概率。文献[12]提出一种新颖的直接后验概率的置信度计算方法,新的方法将 MLP 计算得到的音素类别的后验概率用于置信度的计算,在基于 Lattice 的关键词检测系统框架下获得比 Lattice 后验概率更好的置信度效果。本文拓展和改进了其思想,提出了基于 MLP 帧级子词后验概率的置信度计算方法,即构造子词类别后验概率和子词一致性得分来获取置信度,并将其应用在基于垃圾模型的关键词检测系统框架下。首先通过关键词-补白网络产生候选关键词和补白的识别结果,然后通过时间对齐得到关键词子词的边界信息,利用 MLP 对汉语的声韵母(统称为子词)基元构建声学模型,在关键词子词层将 MLP 输出的各子词类别的后验概率直接用于关键词置信度的计算。实验结果表明,新的方法优于主流的基于似然比检验和在线垃圾得分的方法,具有较好的置信度效果。将提出的置信度方法与主流的似然比置信特征和关键词在线垃圾得分进行线性融合,进一步提高了置信度确认的效果。在微软语料库 500 句测试集中,与主流的方法相比,系统的 EER 相对提高了 11.5%。

## 2 基于似然比检验的置信度方法

基于似然比检验的置信度方法利用假设检验的理论解决语音识别中的置信度计算问题,主要用在基于垃圾模型的关键词检测系统<sup>[13]</sup>中。如图 1 所示,基于垃圾模型的关键词检测系统将关键词模型和垃圾模型组成并行的搜索网络,通过加大对关键词网络的奖赏或者给予垃圾网络一定的惩罚,使得当输入语音流中出现关键词时,关键词分数超过垃圾模组分数,从而检测出关键词。通过调整奖赏或者惩罚的权重,可以改变系统的检出率,同时虚警率也会发生相应的改变,需要采用好的置信度方法对候选的关键词进行确认,在保证高检出率的同时降低系统的虚警率。

基于似然比检验的置信度方法通常是利用语音解码过程的声学模型得分信息,计算正反模型的似然度,即关键词似然得分和非关键词似然得分的比值,当比值超过预先设定的门限值时,认为识别结果正确,否则为错误识别结果。本文的关键词检测系统没有对垃圾词进行专门的建模,关键词网络和垃圾网络共用一套声学模型,采用音节作为声学模型单元时系统检出率

较高,但解码速度很慢,考虑到系统的实时性,本文采用声韵母单元作为声学模型单元。

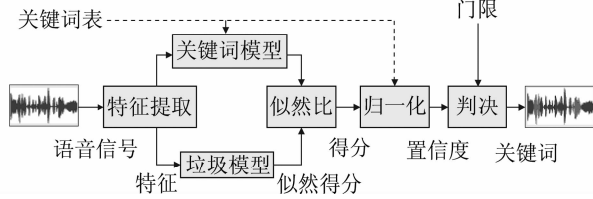


图1 基于垃圾模型的关键词检测系统

Fig. 1 General scheme of keyword spotting system based on garbage model

似然比方法的难点在于反词模型的构建,反词模型越精细,似然比检验的性能就越好,但计算量也越大。明确反词模型的构建具有较好的置信度效果,但需要利用关键词的竞争模型语料通过训练算法训练得到,过程繁琐,不利于任务域的移植。利用声韵母模型集构造反词模型可以省去繁琐的训练过程,本文主要采用这种方法。

关键词在线垃圾得分不同于似然比方法,它的反词部分是由竞争关键词的 HMM 状态得分构成,在关键词的 HMM 状态空间中,取每帧语音状态的似然得分的前  $N$  个值(实验中,  $N = 35$ ) 的均值作为该帧的垃圾得分。该方法反映了关键词和竞争关键词之间的距离,具有一定的抗噪性。

似然比方法和关键词在线垃圾得分都是在 HMM 框架下利用声学模型得分计算得到的,将这两种方法作为基线方法与本文提出的 MLP 框架下基于帧级子词后验概率的置信度方法进行比较,验证新方法的有效性。在此基础上,将两种框架下具有互补性的置信度特征进行融合,进一步提升置信度确认的效果。

### 3 基于 MLP 帧级子词后验概率的置信度方法

MLP 模拟大脑的机理,在机器学习研究中引入了生物神经系统的信息处理机制,使其具有学习和理解能力。MLP 的计算由人工神经元完成,神经元单元之间的权值决定了网络输入输出的任意映射关系。MLP 能很好的描述训练数据中的概率分布特性,当训练数据充足时,MLP 结构可以逼近任意形式的概率分布,能够较为真实的反映语音信号的概率分布。一个标准的 3 层 MLP 网络在语音识别中应用广泛,主要用来估计不同音素类别的后验概率,其结构如图 2 所示。其中,输入层为  $2W+1$  帧长的语音特征,相当于对输入的语音进行加长窗处理,本文实验中  $W$  取 4,输入为 9 帧的

窗序列,输入特征参数采用 39 维的 MFCC 特征,输入层共 351 个节点。输出层为声韵母(Initial/Finals, IF)类别的后验概率。中间层为隐层,包括隐含的神经元单元的个数。

相比 HMM 模型,MLP 网络可以较为准确的描述语音的当前帧和相邻帧的依赖关系,在计算语音当前帧所对应音素类别的后验概率时考虑到了相邻帧的信息。相反,HMM 模型假设每帧语音的声学特征是相互独立的,而声学置信度是局部得分,高度依赖语音的当前帧和相邻帧。因此采用 MLP 网络计算置信度时可以弥补 HMM 模型本身的不足。

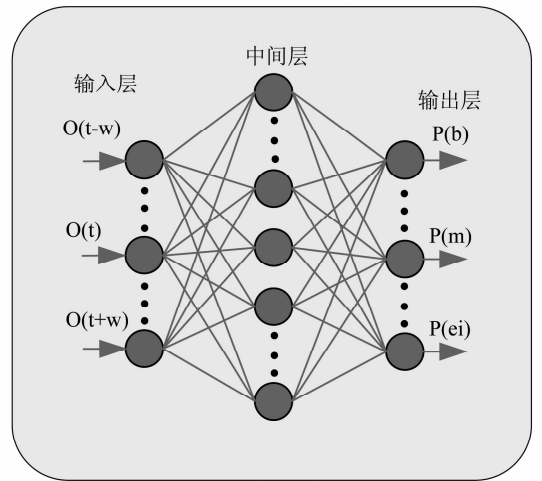


图2 MLP 的结构示意图

Fig. 2 The structure of MLP

本文使用 MLP 估计语音在时间帧  $t$  所对应声韵母类别的后验概率  $P(IF_i | \hat{O}_t)$ ,并提出子词后验概率和子词一致性得分两种方法计算关键词的置信度。其中  $IF_i$  为声韵母类别,实验中采用包括零声母和静音 sil 在内的共 69 个类别; $\hat{O}_t$  为以第  $t$  帧语音为中心的长窗观测序列  $(O_{t-w}, \dots, O_t, \dots, O_{t+w})$ ,  $O_t$  为第  $t$  帧语音所对应的观测序列,为区分  $O_t$ , 记为  $\hat{O}_t$ 。假设关键词检出模块搜索到的关键词  $K_{t_1}^{t_2}$  包括  $M$  个子词  $(IF_1, IF_2, \dots, IF_M)$ , 其中第  $i$  个子词  $IF_i$  的起止时间帧为  $t_{i1}, t_{i2}$ 。通过 MLP 网络的输出可以获得  $t_{i1} \sim t_{i2}$  时间段每帧语音所代表的关键词子词单元  $IF_i$  的后验概率  $P(IF_i | \hat{O}_t)$ , 下面主要给出利用 MLP 输出帧级子词后验概率结果计算关键词置信度的两种不同方法。

#### 3.1 基于子词后验概率的置信度方法

子词后验概率为  $t_{i1}$  到  $t_{i2}$  时间内连续多帧语音  $\hat{O}_{t_{i1}}^{t_{i2}} = \{\hat{O}_{t_{i1}}, \hat{O}_{t_{i1}+1}, \dots, \hat{O}_{t_{i2}}\}$  所对应  $IF_i$  的后验概率  $P(IF_i^{t_{i1}} =$

$IF_i | \hat{\mathbf{O}}_{t_{i1}}^{t_{i2}}$ ),为了便于计算,最容易想到的是直接将多帧语音对应子词类别后验概率进行相乘,即:

$$P(IF_i^{t_{i1}^{t_{i2}}} = IF_i | \hat{\mathbf{O}}_{t_{i1}}^{t_{i2}}) = \prod_{t=t_{i1}}^{t_{i2}} P(IF_i | \hat{\mathbf{O}}_t) \quad (1)$$

但这种方法假设每帧语音对应子词类别的后验概率是独立的,而实际上每帧语音的概率依赖于对应的子词类别。相反,已知关键词子词单元  $IF_i$  的条件下,我们可以合理的假设其所对应的每帧语音之间是相互独立的,即:

$$P(\hat{\mathbf{O}}_{t_{i1}}^{t_{i2}} | IF_i^{t_{i1}^{t_{i2}}} = IF_i) = \prod_{t=t_{i1}}^{t_{i2}} P(\hat{\mathbf{O}}_t | IF_i) \quad (2)$$

则连续多帧语音所对应  $IF_i$  的后验概率为:

$$\begin{aligned} P(IF_i^{t_{i1}^{t_{i2}}} = IF_i | \hat{\mathbf{O}}_{t_{i1}}^{t_{i2}}) &= \frac{P(\hat{\mathbf{O}}_{t_{i1}}^{t_{i2}}, IF_i^{t_{i1}^{t_{i2}}} = IF_i)}{\sum_{IF} P(\hat{\mathbf{O}}_{t_{i1}}^{t_{i2}}, IF_i^{t_{i1}^{t_{i2}}} = IF)} \\ &= \frac{\prod_{t=t_{i1}}^{t_{i2}} P(\hat{\mathbf{O}}_t, IF_i)}{\sum_{IF} \prod_{t=t_{i1}}^{t_{i2}} P(\hat{\mathbf{O}}_t, IF)} = \frac{\prod_{t=t_{i1}}^{t_{i2}} P(IF_i | \hat{\mathbf{O}}_t) P(\hat{\mathbf{O}}_t)}{\sum_{IF} \prod_{t=t_{i1}}^{t_{i2}} P(IF | \hat{\mathbf{O}}_t) P(\hat{\mathbf{O}}_t)} \\ &= \frac{\prod_{t=t_{i1}}^{t_{i2}} P(IF_i | \hat{\mathbf{O}}_t)}{\sum_{IF} \prod_{t=t_{i1}}^{t_{i2}} P(IF | \hat{\mathbf{O}}_t)} \quad (3) \end{aligned}$$

在得到关键词子词的后验概率后,对其求算术均值,就可以得到关键词的置信度,计算公式定义为:

$$CM_{mlp}(K_{t_s}^{t_e}) = P(K_{t_s}^{t_e}) = \frac{1}{M} \sum_{i=1}^M P(IF_i^{t_{i1}^{t_{i2}}} = IF_i | \hat{\mathbf{O}}_{t_{i1}}^{t_{i2}}) \quad (4)$$

### 3.2 子词一致性得分

与子词后验概率的方法不同,子词一致性得分指的是搜索到的关键词所包含的子词单元在时间上与 MLP 输出的子词类别的一致性程度,MLP 输出的子词类别取后验概率最大值所对应的声韵母单元。

假设搜索到的关键词  $K_{t_s}^{t_e}$  的第  $i$  个子词  $IF_i$  的起止时间帧为  $t_{i1}, t_{i2}$ , 对于  $t_{i1} \sim t_{i2}$  时间段的每帧语音,在 MLP 网络的输出结果中找到第  $t$  帧子词类别后验概率最大值所对应的子词单元  $\hat{IF}_t$ , 如果与子词  $IF_i$  相同,该帧语音的一致性得分  $Score$  为 1, 否则为 0。最后,关键词子词  $IF_i$  的置信度通过对所有帧求算术均值获得,如式(6)所示。

$$\hat{IF}_t = \arg \max_{IF} P(IF | \hat{\mathbf{O}}_t) \quad (5)$$

$$CM_{cs}(IF_i) = \frac{1}{t_{i2} - t_{i1} + 1} \sum_{t=t_{i1}}^{t_{i2}} Score(t) \quad (6)$$

其中  $P(IF_i | \hat{\mathbf{O}}_t)$  为每帧语音所代表的关键词子词单元  $IF_i$  的后验概率,  $Score(t)$  为子词的一致性得分,即:

$$Score(t) = \begin{cases} 1, & \text{if } IF_i = \hat{IF}_t \\ 0, & \text{else} \end{cases} \quad (7)$$

在得到关键词子词的后验概率后,就可以计算关键词的置信度,计算公式定义为:

$$CM_{cs}(K_{t_s}^{t_e}) = \frac{1}{M} \sum_{i=1}^M CM_{cs}(IF_i) \quad (8)$$

### 3.3 置信度的融合

在生成式模型 HMM 的框架下,本文采用了主流的基于似然比检验和在线垃圾得分两种置信度方法。在线垃圾得分来自关键词解码空间,在一定程度上体现了关键词和竞争关键词之间的距离,而似然比特征并没有考虑到关键词之间的影响,两者之间具有一定的互补性;在区分性模型 MLP 的框架下,提出了基于子词后验概率和子词一致性得分两种新的置信度计算方法,与主流的方法采用了不同的模型框架,利用了不同的得分信息,这两种模型框架下的置信度方法之间具有一定的互补性。因此本文尝试联合多置信度特征,采用线性加权的方式平衡不同置信度方法的得分,构造一个混合的置信度计算方式,进一步提升置信度确认的效果。

## 4 实验结果

本文实验部分采用微软语料库 Speech Corpora (Version 1.0), 其中用于训练的语料是 100 个年龄在 18-40 岁间的男性录音(大部分在 25 岁以下), 每人大约 200 句话, 总共 19688 句, 454315 个音节。用于测试的语料是 25 个男性录音, 每人 20 句话, 共 500 句, 总时长 44 分 16 秒, 随机挑选 100 个词作为关键词, 其中 70 个两字词, 20 个 3 字词, 10 个 4 字词。各关键词在整个测试集中出现 384 次, 其中 250 句测试语料中不含有关键词。其全部语料在安静办公室环境下录制, 采样频率为 16000Hz, 16bit 量化。

HMM、MLP 模型的语音特征参数均采用 MFCC 参数, 包括语音对数能量、12 维静态 MFCC 参数及其一阶、二阶差分参数, 共计 39 维特征参数。语音信号用 Hamming 窗处理, 帧长 25 毫秒, 帧移 10 毫秒, 语音信号作预加重的系数为 0.97。采用无声调的声韵母作为声学基本单元。

MLP 模型训练时所采用的声韵母单元的时间标注信息采用基于 HMM 的连续语音识别器对训练语料进行时间强制对齐得到, 使用 NICO<sup>[14]</sup> 工具训练神经

网络。

### 4.1 关键词识别系统的性能评价指标

为了验证置信度确认的效果,将识别出的关键词分为正确和错误两大类,然后寻找一个合适的门限,用来决定接受或是拒绝识别的关键词。本文采用检测错误折衷曲线 (Detection Error Tradeoff, DET) 和等错误点 (Equal Error Rate, EER) 评价关键词检测系统的性能。

在关键词检测系统中,漏警率和虚警率分别定义如下:

$$\text{漏警率} = \frac{\text{系统没有检测出的关键词数}}{\text{语音文件中实际出现的关键词数}} \times 100\% \quad (9)$$

$$\text{虚警率} = \frac{\text{识别错误的关键词数}}{\text{语音长度(小时)} \times \text{关键词表大小} \times C} \quad (10)$$

其中  $C$  为常数,作用是使虚警率和漏警率处于同一尺度,物理含义是平均每小时每关键词允许的最大误报数,本实验中取  $C = 10$ 。

DET 曲线定义为不同的置信度判决门限下,系统的漏警率随虚警率的变化趋势,反映着系统的综合性能,在某一特定门限值时,可使得漏警率和虚警率相等,此时的错误率称为系统的等错误率 (EER),反映着系统的工作点情况。

### 4.2 实验结果及分析

HMM 框架下基于似然比检验的置信度方法记为 LLR,关键词在线垃圾得分为 OLG,二者融合的方法记为 LLR+OLG;新提出的 MLP 框架下子词后验概率的置信度方法记为 MLP,子词一致性得分记为 CS。

表 1 对比了 HMM 框架下不同置信度计算方法的 EER。可以看出,似然比方法优于 OLG 特征,联合两种置信特征之后的 EER 下降为 29.5%,这说明两种置信度确实具有一定的互补性。从实验上验证了第 2 节理论分析的正确性。

表 1 HMM 框架下不同置信度方法的 EER 对比

Tab.1 The EER of various confidence measures in the scheme of HMM

Method	LLR	OLG	LLR+OLG
EER	30.6%	31.3%	29.5%

表 2 给出了 MLP 框架下不同置信度计算方法及多置信度联合后的 EER,图 3 为相应的 DET 曲线图。可以看出,本文提出的 MLP 框架下的基于子词类别后验概率和子词一致性得分这两种置信度方法明显优于

传统的方法,较似然比与 OLG 联合特征 (LLR+OLG) 的 EER 分别下降了 1.4% 和 2%,是一种有效的置信度计算方法。与 HMM 模型框架下的置信度特征进行融合之后,性能得到进一步的提升,最优性能的 EER 为 26.1%,与 LLR+OLG 相比,EER 相对提高了 11.5%,说明了这两种不同框架下的置信度特征确实具有一定的互补性。从实验上验证了第 2 节理论分析的正确性。

表 2 MLP 框架下不同置信度方法及多置信度特征联合后的 EER 对比

Tab.2 The EER of various confidence measures in the scheme of MLP and the integrated confidence

Method	MLP	CS	LLR+OLG+MLP	LLR+OLG+CS
EER	28.1%	27.5%	26.9%	26.1%

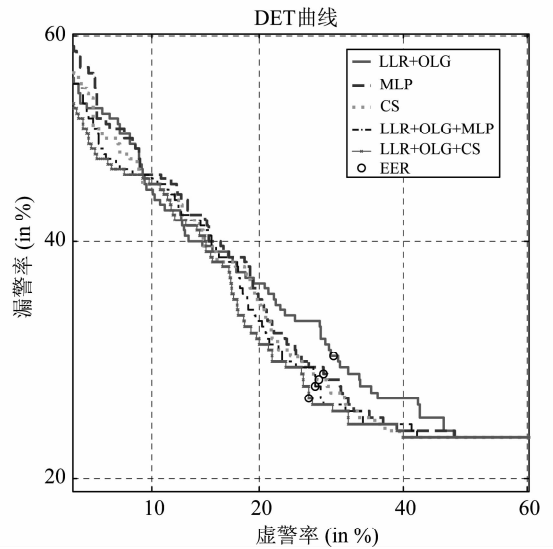


图 3 比较不同关键词置信度计算方法的 DET 曲线  
Fig.3 DET curves using various confidence measures

## 5 结论

本文针对 HMM 模型框架下置信度计算存在的不足,在 MLP 模型框架下提出了基于子词后验概率和子词一致性得分两种新的置信度计算方法,直接将 MLP 输出的每帧语音类别的后验概率用于关键词置信度的计算,关键词检出和置信度确认使用两套不同的模型结构,是两个完全独立的过程,便于融合其他的置信度特征。实验结果表明,MLP 模型框架下的置信度计算方法优于主流的 HMM 模型框架下的方法,两种框架下不同置信度方法融合之后,性能得到进一步提升,相对于主流的计算方法,系统的等错误率 (EER) 相对提高了 11.5%。

## 参考文献

- [1] 王炳锡, 屈丹等. 实用语音识别基础[M]. 北京: 国防工业出版社, 2005.  
Wang Bing-xi, Qu Dan, et al. Practical Fundamentals of Speech Recognition[M]. Beijing. National Defence Industry Press. 2005. (in Chinese)
- [2] Igor Szoke, Petr Schwarz, Pavel Matejka, et al. Phoneme based acoustics keyword spotting in informal continuous speech[C]. In Proc. of RADIOELEKTRONIKA, Brno, Czech Republic, 2005; 302-309.
- [3] D. Vergyri, I. Shafran, A. Stolcke, et al. The SRI/OGI 2006 spoken term detection system[C]. In Proc. of Interspeech, 2007; 2393-2396.
- [4] O. Siohan, B. Ramabhadran, J. Mamou. The IBM 2006 spoken term detection system[C]. In Proc. NIST Spoken Term Detection Evaluation workshop, 2006.
- [5] Jiang Hui. Confidence measures for speech recognition: A survey[J]. Speech Communication, 2005; 455-470.
- [6] 国玉晶, 刘刚, 刘健, 郭军. 基于环境特征的语音识别置信度研究[J]. 清华大学学报(自然科学版), 2009, 49(1); 26-31.  
GUO Yu-jing, LIU Gang, LIU Jian, GUO Jun. Environmental features based confidence measure for speech recognition[J]. Journal of Tsinghua University, 2009, 49(1); 26-31. (in Chinese)
- [7] Jiang Hui. A Dynamic In-Search Data Selection Method With Its Applications to Acoustic Modeling and Utterance Verification[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2005, 13(5); 945-955.
- [8] 孙成立, 刘刚, 郭军. 中文语音确认中子词置信度信能的研究[J]. 中文信息学报, 2008, 22(2); 106-109.  
SUN Cheng-li, LIU Gang, GUO Jun. Investigation of Subwords Confidence Performance in Chinese Speech Verification[J]. Journal of Chinese Information Processing, 2008, 22(2); 106-109. (in Chinese)
- [9] Jie Gao, Qingwei Zhao, Ran Xu and Yonghong Yan. Improved Lattice-based Confidence Measure for Speech Recognition via a Lattice Cut off Procedure[J]. IEEE Computer Science, 2009; 473-476.
- [10] Frank Wessel, Ralf Schluter, Klaus Macherey, et al. Confidence Measures for Large Vocabulary Continuous Speech Recognition [J]. IEEE Transactions on Speech and Audio Processing, 2001, 9(3); 288-298.
- [11] Atsunori Ogawa, Atsushi Nakamura. Discriminative Confidence And Error Cause Estimation For Extended Speech Recognition Function [C]. In: Proceedings of ICASSP, 2010; 4454-4457.
- [12] Dong Wang, Javier Tejedor, Joe Frankel, et al. Posterior-Based Confidence Measures for Spoken Term Detection [C]. In: Proc. of ICASSP, 2009; 4889-4892.
- [13] Igor Szoke, Petr Schwarz, Pavel Matejka, et al. Comparison of keyword spotting approaches for informal continuous speech [C]. In Proc. of Interspeech, 2005; 633-636.
- [14] N. Strom. "The NICO Artificial Neural Network Toolkit". <http://nico.nikkostrom.com>.

## 作者简介



李文昕(1988-),男,河南信阳人,解放军信息工程大学信息工程学院硕士研究生,研究方向:语音关键词识别、置信度计算。E-mail:liwenxin1988@163.com



屈丹(1974-),女,吉林九台人,解放军信息工程大学信息工程学院副教授,硕士生导师,研究方向:语音信号处理、模式识别。E-mail:qudanqudan@sina.com



李弼程(1970-),男,湖南衡南人,解放军信息工程大学信息工程学院教授,博士生导师,研究方向:智能信息处理、语音信号处理。



刘崧(1981-),男,黑龙江林甸县人,杭州恒生数字有限公司,本科,助理工程师,研究方向:信号与信息处理、广播与电视工程。