

文章编号: 1001-0920(2011)05-0777-04

## 一种基于 SVM 的加权似然比融合算法

魏振华, 江雪峰

(华北电力大学 控制与计算机工程学院, 北京 102206)

**摘要:** 提出一种基于改进竞争模型加权似然比融合的方法来进行关键词的语音确认, 并将误差函数用于 SVM 对加权向量和阈值进行优化. 通过构建两个基线系统, 将不同竞争模型的加权方式的比较和 3 种模型的比较作了仿真实验. 所得结果显示, 相对于反词模型和常规竞争模型, 加权似然比融合方法得到的似然比对关键词识别的平均错误率分别降低了 26.53% 和 7.78%.

**关键词:** 置信度测量; 竞争模型; 语音确认

**中图分类号:** TP18

**文献标识码:** A

## Weighted likelihood ratio fusion algorithm based on SVM

WEI Zhen-hua, JIANG Xue-feng

(Department of Control Science and Computer Engineering, North China Electric Power University, Beijing 102206, China. Correspondent: WEI Zhen-hua, E-mail: longerforest@hotmail.com)

**Abstract:** An approach of verification of key words the weighted likelihood ratio(LR) fusion based on the revised rival-model is proposed to be applied to the utterance. Error function is proposed in order to realize the parameter optimization based on the SVM. In the experiment stage, two baseline systems are setup. In the first experiment, the different weight methods based on the rival model are distinguished. In the second experiment, three different models are compared. The experiment result shows that, the LR computed by the weighted likelihood ratio fusion makes a better performance on the keyword spotting, which decreases 26.53% and 7.78% comparing to the normal rival-model and classical anti-model respectively.

**Key words:** confidence measurement; rival model; utterance verification

### 1 引言

近 10 年来, 自动语音识别技术快速发展, 促使自动语音识别系统从实验室迁移到许多实际的服务和应用中. 但是, 在许多自动语音识别应用程序中常需要为验证部分假设识别的正确性提供一种确认机制. 而语音确认正是这样一个过程, 它是用来验证从语音识别器所识别出结果的可靠性. 通常, 定量评分(也称置信度评分)是用来指示每个识别决策的可靠性, 它标识语音识别系统输出的单个单词在置信度评分与给定阈值情况下是正确还是错误<sup>[1]</sup>. 受说话人确认技术的启发, Rahim 等人<sup>[2]</sup>提出利用假设检验来解决置信度计算的问题. 这种使用假设检验的方法解决语音置信度问题被称为似然比(LR)方法. 通常情况下, 似然比是目标假设的 HMM 模型( $H_0$ )似然度与替代假

设模型( $H_1$ )似然度的比值.

替代假设模型  $H_1$  相当复杂. 它包含了目标假设模型空间的补空间, 而且很难用确切的数学建模进行描述. 确认算法的好坏主要取决于对替代假设概率密度函数的拟合能力. 使用由区分性训练过程得到的反词模型来建模替代假设  $H_1$ , 取得了一定的成果, 但不是很理想. 后来, 文献[3]提出了一个“邻近竞争模型”的概念, 并利用它进行语音确认. 这是用一个比较有效的方法来对复杂的反词模型进行建模, 但所有的竞争模型都同等对待, 没有区分不同竞争模型之间对识别的贡献程度.

本文在使用相同的假设模式来研究关键词检测中的语音确认技术之上, 提出一种基于竞争模型的加权似然比融合方法, 并对加权值和阈值在支持向量

收稿日期: 2010-03-08; 修回日期: 2010-06-16.

基金项目: 国家自然科学基金项目(60775058); 哈尔滨市科学研究基金项目(2005AXJ052); 华北电力大学博士科研基金项目(200722019).

作者简介: 魏振华(1973-), 男, 博士, 副教授, 从事语音识别、视频处理等研究; 江雪峰(1986-), 男, 硕士生, 从事语音识别的研究.

机(SVM)上进行了参数优化. 实验结果显示, 基于竞争模型的加权似然比方法取得了较好的实验效果.

## 2 研究背景

关键词检测是语音识别中一门非常具有前瞻性和前途的技术, 它能够识别用户的指令或者要求(不用考虑此要求的确切语法), 只要在他们的话语中包含系统定义的一个或者多个关键字. 但是在一个给定的话语中可能包含字典之内的词(IG)和字典之外的词(OOG). 本文将IG划归成空假设或者目标假设, 将OOG看作冒充者或替代假设. 目标模型 $M^t$ 和替代模型 $M^a$ 是基于隐马尔科夫模型(HMM)训练的, 这里, 似然比被定义为

$$\text{LR}(V|M^t, M^a) = \log \frac{p(V^t|M^t)}{p(V^a|M^a)} \underset{H_1}{\overset{H_0}{>}} \tau. \quad (1)$$

其中:  $H_0$  代表空假设或目标假设,  $V^t$  代表由目标模型产生的特征向量;  $H_1$  代表替代模型,  $V^a$  代表由替代模型产生的特征向量;  $\tau$  是给定的决策阈值.

设目标假设概率 $p(V^t|M^t)$ 是给定识别单元正确解码的假设, 替代假设概率 $p(V^a|M^a)$ 是错误解码的假设. 式(1)表示通过似然比与阈值的比较来接受或者拒绝一个相对于词典之内词的特征向量 $V$ 的假设检验过程.

通过单词错误率(WER)或者单词正确率(WAR)来评估一个ASR系统性能的方法, 一般假设所有的输入单词都来自预先制定的词汇表, 并且所有单词都被识别器解码时标注以相同的置信度. 这样的方法中最常用的是诊断试验特征曲线(ROC曲线)<sup>[4]</sup>. 一个显示由IG错误和OOG错误改变而造成系统响应的有效方法, 是在ROC曲线上以 $X$ 轴显示OOG错误接受率(FAR), 以 $Y$ 轴显示IG错误(ERR-IN). ERR-IN包含了IG错误拒绝(FRR)和IG错误接受(FA-IN)两类错误, 而后者是一个错误的识别假设. 在曲线上不同的点代表不同置信度阈值.

## 3 加权似然比融合

根据Newman-Pearson准则, 一个最优检验是计算在 $H_0$ 和 $H_1$ 两种假设之间的似然比<sup>[4]</sup>. 但是, 替代假设 $H_1$ 范围很广并包含了各种情况. 所以, 似然比方法的难点是如何对 $H_1$ 构建合适的模型. 识别算法主要依赖于替代模型概率密度的拟合能力. 文献[6-7]使用相同的模型(HMM模型)构建 $H_1$ , 再使用反模型构建替代假设模型 $H_1$ , 取得了一定成果, 其中反模型通过一些区分性训练过程获得. 为了得到更好的性能, 文献[3]提出将“竞争模型”的概念用于连续语音识别中的语音确认. “竞争模型”在一定范围内取得了良好的效果, 但也存在一定的问题.

考虑下面的情况:

目标模型是“zi”, 竞争集中极大模型是“ci”,  $V(O) = C$ ;

目标模型是“zi”, 竞争集中极大模型是“za”,  $V(O) = C$ .

语音“ci”和“za”具有相同的似然比分数, 但是很明显“ci”比“za”更为可靠. 因为语音“ci”与“zi”相比于“za”具有更相似的发音. 但又是什么原因使得不同的语音之间具有相同的评分呢? 根据分析发现, 先前的工作中没有考虑不同竞争模型之间的差别, 就像上面的例子一样, 模型“ci”应该比模型“za”拥有更高的分数, 但却给予了相同的对待. 所以, 本文提出一种基于改进竞争模型的加权似然比融合方法, 并用于关键词的检测.

通过竞争模型加权的方式来构造反模型有几种方法: 算法平均数加权方式、调和平均数加权方式以及算术平均数加权方式, 它们相应的公式如下:

$$P(O|\lambda_r) = \sum_{i=0}^N w_i P(O|\lambda_i). \quad (2)$$

其中:  $\lambda_r$  是竞争模型;  $w_i$  是权值并满足条件  $w_i > 0$ ,  $\sum_{i=0}^N w_i = 1$ .

$$P(O|\lambda_r) = (w_1 + w_2 + \dots + w_N) / \sum_{i=0}^N \frac{w_i}{\lambda_i}, \quad (3)$$

$$P(O|\lambda_r) = [P(O|\lambda_1)^{w_1} P(O|\lambda_2)^{w_2} \dots P(O|\lambda_N)^{w_N}]^{1/(w_1+w_2+\dots+w_N)}. \quad (4)$$

使用这些方法来搭建实验平台并作出比较, 旨在找到一种令加权更为优秀的方式. 但是, 在做此工作之前, 还需继续分析上面的公式以找到一种可行的计算方式. 以几何平均数为例:

首先, 将式(6)代入似然比公式(1)得

$$V(O) = \frac{\ln(P(O|\lambda))}{\ln(P(O|\bar{\lambda}))} = \frac{\log \left\{ \left( \frac{P(O|\lambda)}{P(O|\lambda_1)} \right)^{w_1} \dots \left( \frac{P(O|\lambda)}{P(O|\lambda_N)} \right)^{w_N} \right\}^{\frac{1}{w_1+\dots+w_N}}}{\frac{1}{w_1+\dots+w_N} \left( w_1 \log \left( \frac{P(O|\lambda)}{P(O|\lambda_1)} \right) + \dots + w_N \log \left( \frac{P(O|\lambda)}{P(O|\lambda_N)} \right) \right)} = \frac{w^T x}{w_1 + \dots + w_N},$$

其中 $X$ 代表

$$\left[ \log \left( \frac{P(O|\lambda)}{P(O|\lambda_1)} \right), \dots, \log \left( \frac{P(O|\lambda)}{P(O|\lambda_N)} \right) \right].$$

给定阈值 $\tau$ , 并假设 $\theta = (w_1 + w_2 + \dots + w_N)\tau$ , 可以把似然比求解问题变成参数优化问题

$$F(O) = w^T x - \theta. \quad (5)$$

在这里,  $F(O)$  的结果大于0, 识别结果为接受; 否则假

设将被否定.

到此为止,便可给出计算置信度的公式

$$C(W) = \frac{1}{L} \sum_{i=1}^L \frac{1}{1 + \exp(-\rho v_i(o_L))}. \quad (6)$$

其中:  $L$  表示字词的数量, 参数  $\rho$  表示适应参数.

#### 4 基于SVM的参数优化

SVM方法是Vapnik为实现结构化最小风险原则而提出的<sup>[8]</sup>. 它的基本思想是找出一个最好的超平面将两类待分类的样本正确分类, 且使得两类分类数据与超平面之间的距离(分类间距)最大. 设  $x_i (i = 1, 2, \dots, N)$  是训练集  $X$  中的特征向量, 这些向量来自于  $w_1, w_2, \dots, w_N$ , 目的是设计一个超平面, 将所有的训练向量正确分类, 有

$$g(x) = w^T x + b = 0. \quad (7)$$

为了对式(5)相应的权值  $w$  和阈值  $\theta$  进行参数优化, 可构造一个误差函数, 使得这个函数等于样本类的标与分类器对样本类的结果之间的差, 即

$$\sum_{k=1}^n \text{ERR}(x_k, \theta) = \sum_{k=1}^n \left| y - \text{sgn} \left[ \sum_{i=1}^{sv} y_i \alpha_i k(x_i, x_k; \theta) + b \right] \right|. \quad (8)$$

其中:  $y$  表示样本类的标,  $\text{sgn} \left[ \sum_{i=1}^{sv} y_i \alpha_i k(x_i, x_k; \theta) + b \right]$  表示分类器对样本  $x$  的分类,  $x_i$  表示第  $i$  个支持向量,  $y_i$  表示第  $i$  个支持向量的标,  $\alpha_i$  表示支持向量的拉格朗日系数.  $\theta$  为核函数的参数,  $k(x_i, x_k; \theta)$  为核函数, 它可以有几种形式.

参数优化问题是对式(8)求最小值. 将式(8)转化为

$$\arg \min \sum_{k=1}^n (\text{ERR}(x_k, \theta))^2 = \arg \min \sum_{k=1}^n \left| y - \text{sgn} \left[ \sum_{i=1}^{sv} y_i \alpha_i k(x_i, x_k; \theta) + b \right] \right|^2, \quad (9)$$

对其求导, 得

$$\begin{aligned} & -2 \sum_{k=1}^n \left| y - \text{sgn} \left[ \sum_{i=1}^{sv} y_i \alpha_i k(x_i, x_k; \theta) + b \right] \right| \cdot \\ & \frac{\partial \text{sgn} \left[ \sum_{i=1}^{sv} y_i \alpha_i k(x_i, x_k; \theta) + b \right]}{\partial \theta} \cdot \\ & \sum_{i=1}^{sv} y_i \alpha_i \frac{\partial k(x_i, x_k; \theta)}{\partial \theta} = 0. \end{aligned} \quad (10)$$

如果正确分类, 则式(10)左边第1项为0; 否则, 由于右边等于0, 式(10)第1项便可消掉, 于是求核函数参数变为

$$\begin{aligned} & \sum_{k=1}^n \left\{ \frac{\partial \text{sgn} \left[ \sum_{i=1}^{sv} y_i \alpha_i k(x_i, x_k; \theta) + b \right]}{\partial \theta} \right\} \cdot \\ & \sum_{i=1}^{sv} y_i \alpha_i \frac{\partial k(x_i, x_k; \theta)}{\partial \theta} = 0. \end{aligned} \quad (11)$$

对符号函数的求导结果为非零值, 则求核函数参数的方程变为

$$\sum_{k=1}^n \sum_{i=1}^{sv} y_i \alpha_i \frac{\partial k(x_i, x_k; \theta)}{\partial \theta} = 0. \quad (12)$$

以上讨论为了优化权值和阈值, 给出了一个误差函数, 只要使误差函数的值最小, 识别结果更接近于真实的结果. 所以对于本文中改进的竞争集, 可通过SVM训练算法对每个字词求加权向量  $w$  和阈值  $\theta$ . 具体的SVM训练算法如下:

1) 初始化合成权重系数及阈值, 令最易混淆的似然比加权值  $w_i = 1$ , 其余项的似然比加权值都为零, 阈值  $\theta = 0$ ;

2) 使用音素解码器将输出结果与标准答案相比较, 求出误差函数(8)的值;

3) 用SVM训练方法迭代更新式(10)~(12)中各个系数;

4) 重复2)和3), 直到误差函数的值最小为止(收敛).

#### 5 实验及结果分析

为了检验所提出方法的可行性, 本文在中国高技术研究和计划发展的培训和测试数据库上进行实验. 在HMM训练阶段, 采用剑桥大学开发的专门用于建立和处理HMM的实验工具包HTK(HMM Tools Kit).

测试数据库包括27个声母和38个韵母, 每个模型5个状态, 每个状态含有8个高斯混合分量. 语音采用8kHz/16位采样, 帧长25ms, 帧移10ms. 这里的声学特征采用的是39维的MFCC, 其中包含对数能量和12维MFCC参数静态特征以及其一阶、二阶差分特征. 测试系统词表为200个关键字.

在解码过程中实现了一个重要的工作, 即将确认分数存储起来以有效地避免对它们进行重复计算. 这种方法可以显著地提高识别效果.

在建立关键词的基础之上, 对HTK输出的结果(一般为lattice结果)进行剪枝, 因为关键词是所期望得到的最终识别结果. 当一条候选路径到达一个子树末尾, 得到一个关键词假设之后, 对这个关键词进行似然比检验, 从而决定该候选路径的保留与否.

为了验证哪种加权似然比方法具有更好的识别率, 需在基线系统中采用不同的竞争加权方法来实现, 分别是算术平均、几何平均以及调和平均, 3种不同竞争模型加权方法所得的平均错误率(EER)如表1所

表1 不同竞争模型加权方法的EER比较 %

加权方式	算术平均	几何平均	调和平均
平均错误率	29.12	27.21	28.93

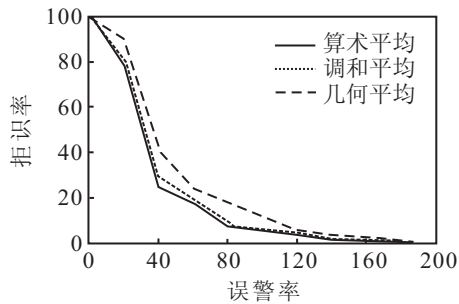


图1 不同竞争模型加权方式的DET曲线

示, 所得的DET曲线如图1所示。

为了评估基于竞争模型上加权似然比融合的语音确认方法的效果, 将它与一般竞争模型和一般反词模型相比较。这里所有的模型都是基于最大似然准则训练的。

在不同的方法下, 对正确识别的词和错误识别的词进行语音确认。表2显示了测试集上的平均错误率。相同的图表也反映在图2中的ROC曲线上。

表2 在测试集上平均错误率的比较 %

测试集	反模型	竞争模型(一般)	竞争模型(修改)
平均错误率	41.12	33.41	30.21

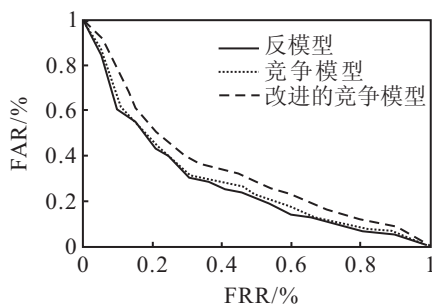


图2 不同模型的ROC曲线

从表1中可以看出, 采用几何平均方式加权竞争模型相比算术平均和调和平均取得了更好的效果, 调和平均略优于算术平均。经分析得出, 对于算术平均方式, 如果目标模型 $\lambda$ 与其所有的竞争模型无关, 或者 $P(O|\lambda)$ 的值很小时, 通常会出现 $P(O|\lambda_r)$ 约等于0的情形, 且在计算置信度时使用的是概率的log对数, 如果它的域出现为0, 则会严重影响置信度的值, 进而对语音识别结果造成扭曲。而调和平均是在算术平均基础之上计算的, 它也会受到相应的影响。所以综合而言, 几何平均方法对竞争模型加权更为有效。

由表2可见, 改进的竞争模型(加权似然比融合)比一般的竞争模型提高了26.53%, 比反模型提高了18.75%。可以看出, 加权竞争模型对语音确认有较大改进。

## 6 结论

本文在研究关键词的语音确认上创新性地提出了修改竞争模型上的加权似然比融合算法。为了在SVM上优化加权值和阈值, 提出了一种基于误差函数并使其训练得到最小值的算法。国家高技术研究和发展计划(NHTRAPC)上的一些预见性结果显示, 修改的竞争模型在语音确认方面优于一般竞争模型和传统的反模型。

## 参考文献(References)

- [1] Hui Jiang, Frank K Soong, Chin-Hui Lee. A data selection strategy for utterance verification in continuous speech recognition[C]. Euro Speech. Scandinavia, 2001: 2573-2576.
- [2] Rafid A Sukkart, Anand R Setlurt, Mazin G Rahim. Utterance verification of keyword string using word based minimum verification error(WB-MVE) training[C]. ICASSP. Atlanta, 1996: 518-521.
- [3] Jiang Hui, Lee C H. A new approach to utterance verification based on neighborhood information in model space[J]. Speech and Audio Processing, 2003, 11(5): 425-434.
- [4] Eduardo Lleida, Richard C Rose. Utterance verification in continuous speech recognition: Decoding and training procedures[J]. Speech and Audio Processing, 2000, 8(2): 126-139.
- [5] Valentin Ion, Reinhold Haeb-Umbach. A novel uncertainty decoding rule with applications to transmission error robust speech recognition[J]. Audio, Speech, and Language Processing, 2008, 16(5): 1047-1060.
- [6] Zhang Shilei, Jiang Danning. Utterance verification using improved confidence measures based on alignment confusion rate in Chinese digits recognition[C]. Acoustics, Speech and Signal Processing. Taipei, 2009: 1309-1312.
- [7] Man-Hung Siu, Brian Mak. Minimization of utterance verification error rate as a constrained optimization problem[J]. Signal Processing Letters, 2006, 13(12): 760-763.
- [8] Ingunn Amdal, Magne Hallstein Johnsen. Log likelihood ratio based annotation verification of a norwegian speech synthesis database[C]. Signal Processing Symposium. Rejkjavik, 2006: 186-189.