第32卷 第22期 计 算 机 工 程 2006年11月
Vol.32 No.22 Computer Engineering November 2006
·人工智能及识别技术· 文章编号：1000—3428(2006)22—0200—03 文献标识码：A 中图分类号：TP301.6

基于过零间隔点的声纹识别技术

周家术，穆 斌
(同济大学软件学院，上海 201804)

摘 要：讨论了现有身份验证技术的状况以及存在的一些问题，叙述了重要的几项生物特征识别技术的优缺点。描述了声纹识别技术广泛涉及的一些声学特性。介绍了基于过零间隔点技术的声纹识别系统和其中一些关键技术和系统的设计流程。对于最重要部分，该文附上了源代码。

关键词：声纹识别；身份验证；Bhattacharyya 距离；数学模型

Speech Recognition Technology Based on Interval Data of Passing Zero

ZHOU Jiashu, MU Bin
(School of Software Engineering, Tongji University, Shanghai 201804)

【Abstract】 This paper discusses the situation of existing identification technology and some problems among them, and explains advantage and disadvantage of some biometrics methods. The paper also introduces some popular acoustics theory involved broadly in speech recognition technology. It introduces speech recognition system based on interval data of passing zero, refers some crucial technology and the design flow of the system. To the most important part, the paper gives the source code.

【Key words】 Speech recognition; Identification validation; Bhattacharyya distance; Mathematics model

1 现有身份验证技术的状况

目前，国内外个人身份的验证、识别主要依靠有效的个人身份证件(如身份证、工作证等)和设定密码等手段来实现。虽然各国都在不断研发各种加密技术，以此来防止信息的被偷窃、伪造，但随着各种伪造、破译技术的发展，这些传统的验证方法已面临着巨大的挑战。因此必须采用更为先进有效的身份验证方法来代替传统的验证手段。目前各国主要研究的生物特征识别技术包括指纹、虹膜、人脸、声纹等以及由几项组成的多生物特征融合。

2 目前声纹识别系统主要用到的声学特征

声纹识别的基本原理是通过分析人的声音波形，从而提取一些有效的特征值，一般来说是根据每人发音的频率不同，通过构造一个有效的数学建模，由计算机对模型和实际输入的语音进行精确匹配，根据匹配结果辨认出说话人是谁。现在声纹识别已经有了一些比较成熟的技术，其中各种声纹识别系统几乎都用到一些常用的声学特征和模型训练技术。声学特征目前主要包括：线性预测系数 LPC，倒谱系数 CEP，Mel 频标倒谱参数(Mel Frequency Cepstrum Coefficient，MFCC)。训练模块一般分为音频特征提取部分和构造数学模型部分。采用上述特征构造数学模型时，现在的系统一般使用 GMM 和 CHMM 的算法。

3 基于过零间隔点的声纹识别系统的关键技术

3.1 声纹识别系统的评估方法

声纹识别系统的一个最重要的评估方法是准确率。准确率是结果正确的身份验证次数跟总的身份验证次数的比率。对于一次判断有下列 4 种情况(假设甲为测试人)：

	判断为甲的声音	判断非甲的声音
甲的声音	判断正确(情况 1)	判断错误(情况 2)
非甲的声音	判断错误(情况 3)	判断正确(情况 4)

由此可得准确率的数学公式如下：

$$准确率_{(precision)} = \frac{情况1的次数 + 情况2的次数}{总的判断次数}$$

所有的声纹识别系统都是追求上式的最高值。

3.2 关键技术

(1)声音波形数据的抽取。根据“贝叶斯假设”，假定组成声音波形的字(即测试人说的话)在确定说话人的身份的作用上是相互独立的。这样就可以根据测试人的任何话进行识别测试，而不是只能针对某一句话。本文是抽取一个波形的足够数据到一个向量，当然此段波形必须包含有测试人的说话信息，而且向量应该足够大(越大精度越高，但会受到时间、存储空间等各种条件制约，本系统的长度为 100 000)，此段波形包含的说话人的话越多则结果越准确，由此得到的向量的数据可以看作说话人说话的频率信息。

(2)特征项的抽取。本系统是基于过零间隔点的声纹识别。过零间隔点特征的提取分为 2 步：

1)计算过零间隔点并进行一些处理。首先找到波形数据的中间值，即在无声的条件下波形的数据值V_m，本系统分析取得的值为V_m=128。对上述取得的数据计算过零间隔点(V_m

作者简介：周家术(1980－)，男，硕士生，主研方向：模式识别，数据挖掘；穆 斌，硕士、副教授

收稿日期：2005-12-13 **E-mail：**jiashu2888@yahoo.com.cn

作为“零点”)，计算公式如下：

$$B_i = \sum_{j=1}^{n} X_j \{\text{如果} V_j - V_m = i \text{则} X_j = 1, \text{否则} X_j = 0\} (1 \le i \le m)$$

其中向量B用来存放过零间隔点，本系统n=100 000，向量B的长度一般为经验值，由于受人发声的频率所限制，当i达到一定的值后，B_i全部为零，设B的长度为m，本系统中m=400(为了提高精度，m也可以根据B_i的最高不为零的值进行动态调整，这时后面计算概率时也应动态调整，计算完的概率还应转换为相同长度，此系统主要验证基于过零间隔点的声纹识别的可行性，因此没有进行此优化)。得到向量B后，可对向量进行适当处理，通过本系统实践分析，此向量的最高的不为零的几项一般是由于噪音产生的，可进行将其值赋为零。

2)求得过零间隔点的概率，为方便声音特征的分析，本系统对过零间隔点要求其概率，计算公式如下：

$$P_i = \frac{B_i * i}{n} (1 \le i \le m)$$

其中 P 为存放过零间隔点概率的向量，向量长度跟 B 相同。现在声音的特征值已经提取出来，即过零间隔点概率向量，下一步就是对该特征值进行分析，建立有效的数学模型，以此来判断测试人的身份。

(3)特征分析及建立数学模型

1)数学期望(均值):它主要计算过零间隔点概率的概率平均值，是描述数据的“集中趋势”的“特征数”。计算公式为

$$E(B) = \sum_{i=1}^{m} P_i * i$$

2)方差：它也是数理统计学中常用的数据处理方法之一。很多情况下只靠对平均数的统计和比较，还不能判别两组数据谁好谁差；这类问题要研究下去，关键是确定这两组数据偏离各自的平均数的大小，此时可用方差来实现比较。方差越大，这组数据就越离散，数据的波动也就越大；方差越小，这组数据就越聚合，数据的波动也就越小。方差是描述数据“离散程度”的“特征数”。计算公式为

$$D(B) = \sum_{i=1}^{m} P_i * i^2 - (E(B))^2$$

通过上述两种特征值，可以对波形数据进行初步的分析。由于这两种特征值各自反映了过零间隔点数据波形的两个不同方面，因此在声纹识别中必须将二者综合考虑，不可偏废。为此必须寻求一种更有效的建模方法，这种建模方法能综合考虑均值和方差。这也是本系统的重点，即采用 Bhattacharyya 距离公式。

3)Bhattacharyya 距离公式

Bhattacharyya 距离在许多关于统计的识别文章中都会涉及到，是计算两个高斯分布的向量的距离的理想方法，特别适用于处理两个高斯分布。经过分析可以得到过零间隔点概率的数据波形呈现高斯分布，因此 Bhattacharyya 距离很适合此系统的身份验证。而且它具有计算简单、扩展性好等一些十分令人满意的特性。在国外，这个特性广泛应用于图像处理方面。它的公式定义为

$$D_{bhatt} = \frac{1}{8}(M_2 - M_1)^T \left[\frac{\sum_1 + \sum_2}{2}\right]^{-1} (M_2 - M_1) + \frac{1}{2}\ln \frac{\left|\frac{\sum_1 + \sum_2}{2}\right|}{\sqrt{|\sum_1||\sum_2|}}$$

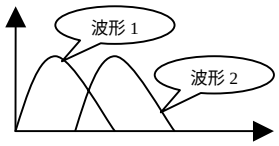
其中：M_i:待比较的向量(本系统为过零间隔点概率的向量)

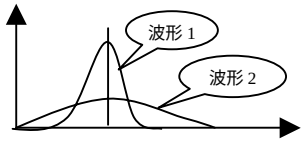
Σ_i: 待比较的向量的协方差(本系统为过零间隔点概率的向量的协方差)

从 Bhattacharyya 公式容易看出，此公式分为 2 部分：

其中第 1 部分：$\frac{1}{8}(M_2 - M_1)^T \left[\frac{\sum_1 + \sum_2}{2}\right]^{-1} (M_2 - M_1)$ 给出了由于在两个向量方法的不同的向量可分离性。第 2 部分：$\frac{1}{2}\ln \frac{\left|\frac{\sum_1 + \sum_2}{2}\right|}{\sqrt{|\sum_1||\sum_2|}}$ 给出了由于在两个向量协方差矩阵的不同的向量可分离性。对于本系统的过零间隔点数据波形，第 1 部分主要是针对图 1 的情况进行区分的；第 2 部分主要针对图 2 的情况进行区分的(这种方式比方差更为有效，精确度更高)。

图 1 Bhattacharyya 公　　**图 2 Bhattacharyya 公式**





式第 1 部分区分的波形　　**第 2 部分有效区分的波形**

因此，将第 1 部分与第 2 部分结合就可有效地对过零间隔点概率数据波形的各种情况进行有效的区分。进一步，对于 Bhattacharyya 距离公式，根据贝叶斯理论，这 2 个过零间隔点概率数据向量距离的最优贝叶斯误差可利用下式计算：

$$\varepsilon \le \sqrt{P_1 P_2} e^{(-D_{bhat})}$$

显然当 P1=P2=0.5 时误差概率取得最大值，即此时可得

$$\varepsilon_{bhat} = 0.5 e^{(-D_{bhat})}$$

从上述中可以看出，用 Bhattacharyya 距离公式的优点如下：(1)非常容易计算；(2)通过指定误差上限而不是精确的解决方法，这就提供了两个向量的“光滑”的距离。这对于我们的需求更适合，因为一般不会相信我们的数据是非常标准的分布。

4 阈值的确定

计算得到 Bhattacharyya 距离的值后，需要根据设定的阈值来判断是否通过身份验证。但是，阈值的确定是十分困难的，理论上，没有很好的解决方法，一般采用预定初始值，然后给出测试数据，再根据测试数据的准确程度调整初始值，这样有 2 个缺点：(1)初始值的确定不容易，完全是根据经验或简单的测试而定；(2)调整的幅度无法确定，只能不断地测试，不断地调整，大大增加了工作量。

5 基于过零间隔点的声纹识别系统的实现

结合了上述的关键技术，实现的基于过零间隔点声纹识别系统的系统结构如图 3 所示。

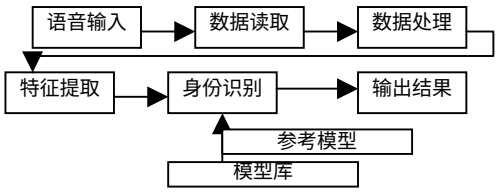


图 3 基于过零间隔点声纹识别系统的系统结构

下面是基于过零间隔点声纹识别系统的详细设计过程及计算 Bhattacharyya 距离公式的源代码：

(1)读取语音的波形数据到一个向量中。本系统存放在一个数组中，数组的大小为 100 000，读取声音文件(本系统只识别.wav 文件)中连续的 100 000 个数据到数组中(本系统的测试数据能保证填满数组的 100 000 数据)。

(2)求得过零间隔点：将上述得到的声音数据利用前面所讲的计算过零间隔点的公式进行计算，以无声时提取的声音数据作为零点，

然后计算上述数据，得到过零间隔的次数并存放在另外一个数组中，经过测试后本系统中此数组的大小为 400。

(3)求得过零间隔点的概率：上一步得到过零间隔点数组，同时可知数据的总量为 100 000，因此根据过零间隔点概率的计算公式计算可得到过零间隔点的概率向量，本系统将此向量存放在另外一个数组中，显然数组的大小也为 400。

(4)提取数据：从模型库中提取要核对的数据。

(5)身份识别：通过上述步骤，可得到模型数据和测试人的过零间隔点概率数据。根据 Bhattacharyya 距离公式进行计算。计算出结果后将结果与阈值相比较即可得出验证结果，如果计算的结果小于先前设定的阈值，则可以判断是同一个人，即通过验证，否则拒绝通过。Bhattacharyya 距离的 C++语言实现方法如下：

```
double CalDistance(CVector& CV1,CVector& CV2, CMatrix& CM1, CMatrix CM2)
{CVector CVDif(CV1-CV2);
    CMatrix CMAdd((CM1+CM2)/2);
    double dAddDeter = fabs(CMatrix(CMAdd*10).GetDeterminant ());
    if(!CMAdd.Invert())
      return 0;
    CMatrix CMDis
    (CMatrix(CVDif)*CMAdd*CMatrix(CVDif,FALSE) *0.125);
    double    Dis = CMDis[0][0];
    double dTe1 = fabs(CMatrix(CM1*10).GetDeterminant());
    double dTe2 = fabs(CMatrix(CM2*10).GetDeterminant());
    double Cov = Dis;
    Dis+=0.5*(log(dAddDeter)-0.5*(log(dTe1)+log(dTe2)));
    Cov = Dis-Cov;
    return Dis; }
```

(6)输出结果。输出是否验证通过。

6 总结

本文提出了一种基于过零间隔点的声纹识别技术，并对当前的身份鉴别技术进行了比较，同时介绍了声纹识别系统广泛涉及到的声学特征。对于本系统的最重要的 Bhattacharyya 计算公式还附带了实现的源代码。经过测试，这种方法的准确率在 93%以上。由此说明，根据过零间隔点进行声纹识别方法是可行的。由于本技术没有涉及一些其它的优化，例如噪音的有效去除，以及一些模糊算法应用等，因此准确率没有达到很高，继续通过别的一些优化方式，可以对声纹识别取得理想的效果。

参考文献

1 金晓伟. 指纹识别在银行业中的应用[J]. 计算机安全, 2005, (2).

2 周静芳. 基于高斯语音滤波的稳健文本无关说话人识别[J]. 计算机工程, 2005, 31(2).

3 Fukunaga K. Introduction to Statistical Pattern Recognition(2nd Edition)[Z]. Academic Press, Inc., 1990.

4 Barnard E, Cole R A, Fanty M, et al. Real-world Speech Recognition with Neural Networks[Z]. Hillsdale, New Jersey: Lawrence Erlbaum Assoc., 1995.

5 Jelinek F. Continuous Speech Recognition by Statistical Methods[J]. Proceedings of the IEEE, 1976, 64(4): 532-556.

(上接第 193 页)

com.cn)合作采用 WOKPS 构建一个面向 Web 的健康咨询平台。该平台将分散在 Web 中与不同主题相关的健康咨询专家系统联结起来，对用户全面的、详细的体检报告进行健康状况评测并给出恰当的建议。用户可以登录到该平台，选择测试的种类，通过网页输入自己的各项指标，并显示评测结果，如图 3 所示。

图 3 面向 Web 的健康咨询平台的用户界面

当用户选定测试的种类并输入各项指标后，协调器查看目录管理中注册的 KBS 的功能描述和输入要求，选定相应的一个或多个 KBS，规划这些 KBS 的执行顺序，并为 KBS 的运行设置参数，然后协调器将执行方案包装为消息发给 MTS。这条消息包括消息的来源(用户和协调器)、消息的种类(执行方案)、消息的接收者(KUPMS)和消息的内容(多个 KBS 的各种信息)。当 KUPMS 接收到该消息之后，根据消息内容中描述的信息，访问各个 KBS 的当前状态，并为每一个 KBS 提供消息处理的接口，将其封装为知识处理单元，为单元的运行分配资源和输入参数，并启动知识处理系统进行推理。用户可以通过控制器查看当前任务的分解方式、各子任务的执行状况以及每个 KBS 的推理过程和结果。

5 结论

WWW 是一个由互联资源构成的网络化生长的信息空间，是人们发布信息、获得信息、取得服务的重要渠道。WOKPS 是一个分布式的基于 Web 知识处理平台，它将分布在 Web 中多个 KBS 联结起来对复杂问题进行求解，为人们提供智能的服务。该平台能确保动态地处理新的成员，且对本地数据和程序没有任何改变，并允许独立的资源所有者继续对自己的资源进行核心的控制。WOKPS 将 Web 技术和面向对象知识处理系统(OKPS)相结合，为多个 KBS 的协作问题求解提供了动态的、分布的、广泛共享的知识处理平台。

参考文献

1 史忠植, 董明楷, 蒋运承等. 智能互联网[J]. 计算机科学, 2003, 30(9): 1-4.

2 史忠植. 智能主体及其应用[M]. 北京: 科学出版社, 2000.

3 史忠植. 知识工程[M]. 北京: 清华大学出版社, 1988.

4 史忠植. 高级人工智能[M]. 北京: 科学出版社, 1998.

5 OKPS[Z]. http://www.intsci.ac.cn/research/okps.html.

6 Essbase. Essbase Web Gateway[Z]. http://www.arborsoft.com.

7 Purchasing Analysis[Z]. http://www.boozallen.com.