



(12) 发明专利

(10) 授权公告号 CN 109473107 B

(45) 授权公告日 2020.12.22

(21) 申请号 201811468428.3

G10L 17/02 (2013.01)

(22) 申请日 2018.12.03

G10L 17/24 (2013.01)

(65) 同一申请的已公布的文献号  
申请公布号 CN 109473107 A

(56) 对比文件

CN 106057206 A, 2016.10.26

CN 106098068 A, 2016.11.09

(43) 申请公布日 2019.03.15

CN 108446638 A, 2018.08.24

(73) 专利权人 厦门快商通信息技术有限公司  
地址 361007 福建省厦门市思明区嘉禾路  
267号9楼902室B区

CN 106531171 A, 2017.03.22

CN 105913850 A, 2016.08.31

US 2007294083 A1, 2007.12.20

(72) 发明人 洪国强 肖龙源 蔡振华 李稀敏  
刘晓葳 谭玉坤

审查员 陈成

(74) 专利代理机构 厦门仕诚联合知识产权代理  
事务所(普通合伙) 35227  
代理人 乐珠秀

(51) Int. Cl.

G10L 17/08 (2013.01)

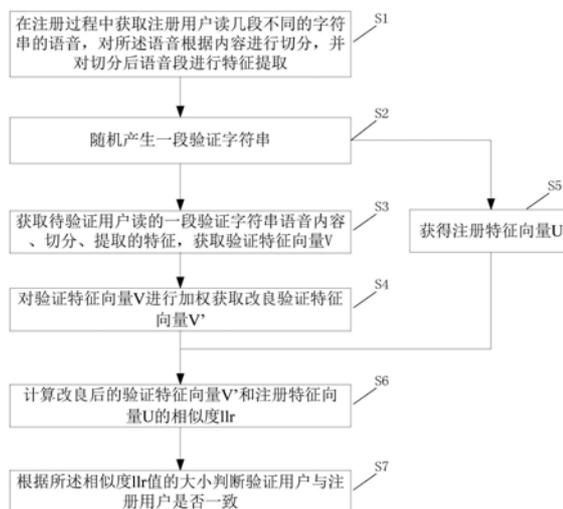
权利要求书3页 说明书8页 附图1页

(54) 发明名称

一种文本半相关的声纹识别方法及系统

(57) 摘要

本发明公开了一种文本半相关的声纹识别方法及系统,所述方法包括:在注册过程中获取注册用户读几段不同的字符串的语音,对所述语音根据内容进行切分,并对切分后语音段进行特征提取;随机产生一段验证字符串,获取验证特征向量V,对验证特征向量V进行加权获取改良验证特征向量V',根据所述一段验证字符串对应的字符获取获得注册特征向量U,计算改良后的验证特征向量V'和注册特征向量U的相似度llr判断验证用户与注册用户是否一致。采用本发明的声纹识别方法及系统具备简单可靠安全,同时充分考虑了各数字的不同识别效果,大大提高声纹识别精度。



1. 一种文本半相关的声纹识别方法,其特征在于,所述方法包括如下:

步骤一,在注册过程中获取注册用户读几段不同的字符串的语音,对所述语音根据内容进行切分,并对切分后语音段进行特征提取;所述切分是将语音内容的每个字符切成一语音段;

步骤二,随机产生一段验证字符串,所述一段验证字符串包含 $n$ 个字符,并获取待验证用户读的一段验证字符串语音内容,并对一段验证字符串语音内容进行切分,分别对每个字符提取的特征,并将提取的特征依次串起来获取验证特征向量 $V$ ,对验证特征向量 $V$ 进行加权获取改良验证特征向量 $V'$ ,其中, $V=[v_1, v_2, \dots, v_n]$ , $V'=[C_1v_1, C_2v_2, \dots, C_nv_n]$ , $v_i$ 为所述一段验证字符串的第 $i$ 个字符的特征, $C_i$ 为所述一段验证字符串的第 $i$ 个字符的特征对应的权值,

其中 $i, n$ 均为正整数, $n \geq 2, 1 \leq i \leq n$ ;

步骤三,根据所述一段验证字符串对应的字符获取注册用户在注册过程中对应的特征,并依次即将所述一段验证字符串对应的字符特征串起来获得注册特征向量 $U, U=[u_1, u_2, \dots, u_n]$ ,其中, $u_i$ 为所述一段验证字符串的第 $i$ 个字符对应的注册特征;

步骤四,计算改良后的验证特征向量 $V'$ 和注册特征向量 $U$ 的相似度 $llr$ ;

步骤五,根据所述相似度 $llr$ 值的大小判断验证用户与注册用户是否一致。

2. 根据权利要求1所述的文本半相关的声纹识别方法,其特征在于,

所述字符串为数字串,所述一段验证字符串包含 $n$ 个数字;

步骤一,在注册过程中获取注册用户读几段不同的数字串的语音,对所述语音根据内容进行切分,并对切分后语音段进行特征提取;所述切分是将语音内容的每个数字切成一语音段;

步骤二,随机产生一段验证数字串,所述一段验证数字串包含 $n$ 个数字,并获取待验证用户读的一段验证数字串语音内容,并对一段验证数字串语音内容进行切分,分别对每个数字提取的特征,并将提取的特征依次串起来获取验证特征向量 $V$ ,对验证特征向量 $V$ 进行加权获取改良验证特征向量 $V'$ ,其中, $V=[v_1, v_2, \dots, v_n]$ , $V'=[C_1v_1, C_2v_2, \dots, C_nv_n]$ , $v_i$ 为所述一段验证数字串的第 $i$ 个数字验证特征, $C_i$ 为所述一段验证数字串的第 $i$ 个数字验证特征对应的权值;

步骤三,根据所述一段验证数字串对应的数字获取注册用户在注册过程中对应的特征,并依次即将所述一段验证数字串对应的数字特征串起来获得注册特征向量 $U, U=[u_1, u_2, \dots, u_n]$ ,其中, $u_i$ 为所述一段验证数字串的第 $i$ 个数字对应的注册特征。

3. 根据权利要求2所述的文本半相关的声纹识别方法,其特征在于,

$C_i$ 为误识率为 $f$ 的情况下,所述一段验证数字串的第 $i$ 个数字单独用于声纹识别的识别率 $r$ 作为所述一段验证数字串的第 $i$ 个数字的验证特征对应的权值,所述一段验证数字串的第 $i$ 个数字为数字“X”,所述识别率 $r$ 的获取过程包括:

(1) 获取 $m$ 个人的数字“X”的语音段,每个人至少包含两段语音,分别提取各语音段的特征向量;

(2) 构建一定数量的正负对,两个特征为同一人即正对,两个特征为不同人为负对;

正对是指注册过程输出的特征与验证过程输出的特征一致,负对是指注册过程输出的特征与验证过程输出的特征不一致,特征是从语音中提取的,如果配对的两段语音不是同

一人的即为负对,为同一人的即为正对;

(3) 分别对一定数量的正负对的特征向量进行打分,其中,正对的特征向量打分大于 $t$ 的数量为 $Tg$ ,正对的特征向量打分小于 $t$ 的数量为 $Tl$ ;负对的特征向量打分大于 $t$ 的数量为 $Fg$ ,负对的特征向量小于 $t$ 的数量为 $Fl$ ,计算误识率为 $f = Fl / (Fg + Fl)$ ;

所述打分是注册特征向量、验证特征向量进行相似度进行计算;

(4) 识别率为 $r = Tl / (Tg + Tl)$ ,其中,所述一段验证数字串的第 $i$ 个数字为数字“X”的 $C_i$ 值即为确定 $f$ 值时的识别率 $r$ 值。

4. 根据权利要求3所述的文本半相关的声纹识别方法,其特征在于,误识率的 $f$ 值可根据场景需求行设定。

5. 根据权利要求3所述的文本半相关的声纹识别方法,其特征在于,构建一定数量的正负对,其中一定数量为不低10000。

6. 根据权利要求1或2所述的文本半相关的声纹识别方法,其特征在于,在步骤四中,通过余弦距离算法获取改良后的验证特征向量 $V'$ 和注册特征向量 $U$ 的相似度 $llr$ 。

7. 一种文本半相关的声纹识别系统,其特征在于,包括:

注册模块,用于提供几段不同的字符串供注册用户读取,获取注册用户读几段不同的字符串的语音,用于对语音进行切分后语音段进行特征提取;所述切分是将语音内容的每个字符切成一语音段;

验证模块,用于随机产生一段验证字符串,所述一段验证字符串包含 $n$ 个字符;

验证特征模块,用于获取待验证用户读的一段验证字符串语音内容,并对一段验证字符串语音内容进行切分,分别对每个字符提取的特征,并将提取的特征依次串起来获取验证特征向量 $V$ , $V = [v_1, v_2, \dots, v_n]$ , $v_i$ 为所述一段验证字符串的第 $i$ 个字符的特征, $v_1$ 为所述一段验证字符串的第1个字符的特征, $v_2$ 为所述一段验证字符串的第2个字符的特征, $v_n$ 为所述一段验证字符串的第 $n$ 个字符的特征,其中 $i$ 、 $n$ 均为正整数, $n \geq 2$ , $1 \leq i \leq n$ ;

改良模块,用于对验证特征向量 $V$ 进行加权获取改良验证特征向量 $V'$ , $V' = [C_1v_1, C_2v_2, \dots, C_nv_n]$ , $C_i$ 为所述一段验证字符串的第 $i$ 个字符的特征对应的权值;

注册特征模块,用于根据所述验证模块产生的所述一段验证字符串对应的字符获取注册用户在注册过程中对应的特征,并依次即将所述一段验证字符串对应的特征串起来获得注册特征向量 $U$ , $U = [u_1, u_2, \dots, u_n]$ ,其中, $u_i$ 为所述一段验证字符串的第 $i$ 个字符对应的注册特征;

判断模块,用于计算改良后的验证特征向量 $V'$ 和注册特征向量 $U$ 的相似度 $llr$ ,并根据所述相似度 $llr$ 值的大小判断验证用户与注册用户是否一致。

8. 根据权利要求7所述的文本半相关的声纹识别系统,其特征在于,

所述注册模块提供几段不同的数字串供注册用户读取,所述几段不同的数字串的语音为几段不同的数字串的语音,所述切分是将语音内容的每个数字切成一语音段;

所述验证模块随机产生一段验证数字串,所述一段验证数字串包含 $n$ 个数字;

所述验证特征模块获取待验证用户读的一段验证数字串语音内容,并对一段验证数字串语音内容进行切分,分别对每个数字提取的特征,并将提取的特征依次串起来获取验证特征向量 $V$ , $V = [v_1, v_2, \dots, v_n]$ , $v_i$ 为所述一段验证数字串的第 $i$ 个数字的特征;

改良模块,用于对验证特征向量 $V$ 进行加权获取改良验证特征向量 $V'$ , $V' = [C_1v_1, C_2v_2, \dots, C_nv_n]$ , $C_i$ 为所述一段验证数字串的第 $i$ 个数字的特征对应的权值;

注册特征模块,用于根据所述验证模块产生的所述一段验证数字串对应的数字获取注册用户注册过程中对应的特征,并依次即将所述一段验证数字串对应的特征串起来获得注册特征向量 $U$ , $U = [u_1, u_2, \dots, u_n]$ ,其中, $u_i$ 为所述一段验证数字串的第 $i$ 个数字对应的注册特征。

9. 根据权利要求8所述的文本半相关的声纹识别系统,其特征在于,

改良模块中的 $C_i$ 为所述一段验证数字串的第 $i$ 个数字的特征对应的权值;误识率为 $f$ 的情况下,所述一段验证数字串的第 $i$ 个数字单独用于声纹识别的识别率 $r$ 作为所述一段验证数字串的第 $i$ 个数字的验证特征对应的权值,所述一段验证数字串的第 $i$ 个数字为数字“X”,改良模块中的获取识别率 $r$ 包括:

获取 $m$ 个人的数字“X”的语音段,每个人至少包含两段语音,分别提取各语音段的特征向量;

构建一定数量的正负对,两个特征为同一人即正对,两个特征为不同人为负对;

正对是指注册过程输出的特征与验证过程输出的特征一致,负对是指注册过程输出的特征与验证过程输出的特征不一致,特征是从语音中提取的,如果配对的两段语音不是同一人的即为负对,为同一人的即为正对;

分别对一定数量的正负对的特征向量进行打分,其中,正对的特征向量打分大于 $t$ 的数量为 $T_g$ ,正对的特征向量打分小于 $t$ 的数量为 $T_1$ ;负对的特征向量打分大于 $t$ 的数量为 $F_g$ ,负对的特征向量小于 $t$ 的数量为 $F_1$ ,计算误识率为 $f = F_1 / (F_g + F_1)$ ;

所述打分是注册特征向量、验证特征向量进行相似度进行计算;

识别率为 $r = T_1 / (T_g + T_1)$ ,其中,所述一段验证数字串的第 $i$ 个数字为数字“X”的 $C_i$ 值即为确定 $f$ 值时的识别率 $r$ 值。

10. 根据权利要求9所述的文本半相关的声纹识别系统,其特征在于,

构建一定数量的正负对,其中一定数量为不低10000;误识率的 $f$ 值可根据场景需求行设定。

## 一种文本半相关的声纹识别方法及系统

### 技术领域

[0001] 本发明涉及声纹识别技术领域,具体涉及与一串随机动态数字相关的一种文本半相关的声纹识别方法。

### 背景技术

[0002] 在现有声纹识别中主要有两种声纹识别方式,分别为固定文本(text-dependent)声纹识别和文本无关(text-independent)声纹识别,在实际应用中,文本无关声纹识别需要很长的语音才能达到理想的识别效果,而固定文本声纹识别只要很短的一句话就能有很好的识别效果。但在实际应用中固定文本声纹识别方式由于文本内容固定,难以防止录音等情况发生,固定文本声纹识别方式存在较大的安全问题。

[0003] 在声纹识别的实际应用中以动态数字串最为常见,其在注册时会给出几个随机数字串,而在验证时给出一个随机数字串用于识别,具体包含:

[0004] (1)注册步骤:获取说话人读几段不同的数字串的语音,对所述语音根据内容进行切分,即将语音内容的每个数字切成一段,并对切分后语音段进行特征提取,

[0005] (2)验证步骤:将说话人读的一段数字串语音进行切分,分别提取特征,然后将提取的特征依次串起来获取验证特征向量V;

[0006] (3)相似度判断步骤:将注册获取的不同数字的特征,根据验证给的数字串依次串起来获得注册特征向量U,通过cosine方式计算V和U的相似度。

[0007] 上述声纹识别方法未考虑到各个数字的识别效果,声纹识别精度较低。

### 发明内容

[0008] 本发明的目的在于克服现有技术问题,提出一种即提高声纹识别精度,又可以从较短语音进行声纹识别,同时可以防止录音等情况产生,确保声纹识别应用安全的文本半相关的声纹识别方法及系统,

[0009] 为达成上述目的,本发明提供了文本半相关的声纹识别方法,所述方法包括如下:

[0010] 步骤一,在注册过程中获取注册用户读几段不同的字符串的语音,对所述语音根据内容进行切分,并对切分后语音段进行特征提取;所述切分是将语音内容的每个字符切成一语音段;

[0011] 步骤二,随机产生一段验证字符串,所述一段验证字符串包含n个字符,并获取待验证用户读的一段验证字符串语音内容,并对一段验证字符串语音内容进行切分,分别对每个字符提取的特征,并将提取的特征依次串起来获取验证特征向量V,对验证特征向量V进行加权获取改良验证特征向量V',其中, $V = [v_1, v_2, \dots, v_n]$ ,  $V' = [C_1v_1, C_2v_2, \dots, C_nv_n]$ ,  $v_i$ 为所述一段验证字符串的第i个字符的特征,  $C_i$ 为所述一段验证字符串的第i个字符的特征对应的权值,

[0012] 其中i、n均为正整数,  $n \geq 2, 1 \leq i \leq n$ ;

[0013] 步骤三,根据所述一段验证字符串对应的字符获取注册用户在注册过程中对应的

特征,并依次即将所述一段验证字符串对应的字符特征串起来获得注册特征向量 $U$ , $U=[u_1, u_2, \dots, u_n]$ ,其中, $u_i$ 为所述一段验证字符串的第 $i$ 个字符对应的注册特征;

[0014] 步骤四,计算改良后的验证特征向量 $V'$ 和注册特征向量 $U$ 的相似度 $llr$ ;

[0015] 步骤五,根据所述相似度 $llr$ 值的大小判断验证用户与注册用户是否一致。

[0016] 可选地,所述字符串为数字串,所述一段验证数字串包含 $n$ 个数字;

[0017] 步骤一,在注册过程中获取注册用户读几段不同的数字串的语音,对所述语音根据内容进行切分,并对切分后语音段进行特征提取;所述切分是将语音内容的每个数字切成一语音段;

[0018] 步骤二,随机产生一段验证数字串,所述一段验证数字串包含 $n$ 个数字,并获取待验证用户读的一段验证数字串语音内容,并对一段验证数字串语音内容进行切分,分别对每个数字提取的特征,并将提取的特征依次串起来获取验证特征向量 $V$ ,对验证特征向量 $V$ 进行加权获取改良验证特征向量 $V'$ ,其中, $V=[v_1, v_2, \dots, v_n]$ , $V'=[C_1v_1, C_2v_2, \dots, C_nv_n]$ , $v_i$ 为所述一段验证数字串的第 $i$ 个数字验证特征, $C_i$ 为所述一段验证数字串的第 $i$ 个数字验证特征对应的权值;

[0019] 步骤三,根据所述一段验证数字串对应的数字获取注册用户在注册过程中对应的特征,并依次即将所述一段验证数字串对应的数字特征串起来获得注册特征向量 $U$ , $U=[u_1, u_2, \dots, u_n]$ ,其中, $u_i$ 为所述一段验证数字串的第 $i$ 个数字对应的注册特征。

[0020] 进一步的, $C_i$ 为误识率为 $f$ 的情况下,所述一段验证数字串的第 $i$ 个数字单独用于声纹识别的识别率 $r$ 作为所述一段验证数字串的第 $i$ 个数字的验证特征对应的权值,所述一段验证数字串的第 $i$ 个数字为数字“X”,所述识别率 $r$ 的获取过程包括:

[0021] (1) 获取 $m$ 个人的数字“X”的语音段,每个人至少包含两段语音,分别提取各语音段的特征向量;

[0022] (2) 构建一定数量的正负对,两个特征为同一人即正对,两个特征为不同人为负对;

[0023] 正对是指注册过程输出的特征与验证过程输出的特征一致,负对是指注册过程输出的特征与验证过程输出的特征不一致,特征是从语音中提取的,如果配对的两段语音不是同一人的即为负对,为同一人的即为正对;

[0024] (3) 分别对一定数量的正负对的特征向量进行打分,其中,正对的特征向量打分大于 $t$ 的数量为 $T_g$ ,正对的特征向量打分小于 $t$ 的数量为 $T_1$ ;负对的特征向量打分大于 $t$ 的数量为 $F_g$ ,负对的特征向量小于 $t$ 的数量为 $F_1$ ,计算误识率为 $f=F_1/(F_g+F_1)$ ;

[0025] 所述打分是注册特征向量、验证特征向量进行相似度进行计算;

[0026] (4) 识别率为 $r=T_1/(T_g+T_1)$ ,其中,所述一段验证数字串的第 $i$ 个数字为数字“X”的 $C_i$ 值即为确定 $f$ 值时的识别率 $r$ 值。

[0027] 进一步的,需要说明的是,误识率的 $f$ 值可根据场景需求行设定。

[0028] 可选地,构建一定数量的正负对,其中一定数量为不低10000。

[0029] 进一步的,在本发明优选实施例中,在步骤四中,通过余弦距离算法获取改良后的验证特征向量 $V'$ 和注册特征向量 $U$ 的相似度 $llr$ 。

[0030] 与上述文本半相关的声纹识别方法相对应的,本发明还提供了一种文本半相关的声纹识别系统,包括:

[0031] 注册模块,用于提供几段不同的字符串供注册用户读取,获取注册用户读几段不同的字符串的语音,用于对语音进行切分后语音段进行特征提取;所述切分是将语音内容的每个字符切成一语音段;

[0032] 验证模块,用于随机产生一段验证字符串,所述一段验证字符串包含 $n$ 个字符;

[0033] 验证特征模块,用于获取待验证用户读的一段验证字符串语音内容,并对一段验证字符串语音内容进行切分,分别对每个字符提取的特征,并将提取的特征依次串起来获取验证特征向量 $V$ , $V=[v_1, v_2, \dots, v_n]$ , $v_i$ 为所述一段验证字符串的第 $i$ 个字符的特征, $v_1$ 为所述一段验证字符串的第1个字符的特征, $v_2$ 为所述一段验证字符串的第2个字符的特征, $v_n$ 为所述一段验证字符串的第 $n$ 个字符的特征,其中 $i, n$ 均为正整数, $n \geq 2, 1 \leq i \leq n$ ;

[0034] 改良模块,用于对验证特征向量 $V$ 进行加权获取改良验证特征向量 $V'$ , $V'=[C_1v_1, C_2v_2, \dots, C_nv_n]$ , $C_i$ 为所述一段验证字符串的第 $i$ 个字符的特征对应的权值;

[0035] 注册特征模块,用于根据所述验证模块产生的所述一段验证字符串对应的字符获取注册用户在注册过程中对应的特征,并依次即将所述一段验证字符串对应的特征串起来获得注册特征向量 $U$ , $U=[u_1, u_2, \dots, u_n]$ ,其中, $u_i$ 为所述一段验证字符串的第 $i$ 个字符对应的注册特征;

[0036] 判断模块,用于计算改良后的验证特征向量 $V'$ 和注册特征向量 $U$ 的相似度 $llr$ ,并根据所述相似度 $llr$ 值的大小判断验证用户与注册用户是否一致。

[0037] 可选的,上述文本半相关的声纹识别系统中,所述注册模块提供几段不同的数字串供注册用户读取,所述几段不同的数字串的语音为几段不同的数字串的语音,所述切分是将语音内容的每个数字切成一语音段;

[0038] 所述验证模块随机产生一段验证数字串,所述一段验证数字串包含 $n$ 个数字;

[0039] 所述验证特征模块获取待验证用户读的一段验证数字串语音内容,并对一段验证数字串语音内容进行切分,分别对每个数字提取的特征,并将提取的特征依次串起来获取验证特征向量 $V$ , $V=[v_1, v_2, \dots, v_n]$ , $v_i$ 为所述一段验证数字串的第 $i$ 个数字的特征;

[0040] 改良模块,用于对验证特征向量 $V$ 进行加权获取改良验证特征向量 $V'$ , $V'=[C_1v_1, C_2v_2, \dots, C_nv_n]$ , $C_i$ 为所述一段验证数字串的第 $i$ 个数字的特征对应的权值;

[0041] 注册特征模块,用于根据所述验证模块产生的所述一段验证数字串对应的数字获取注册用户在注册过程中对应的特征,并依次即将所述一段验证数字串对应的特征串起来获得注册特征向量 $U$ , $U=[u_1, u_2, \dots, u_n]$ ,其中, $u_i$ 为所述一段验证数字串的第 $i$ 个数字对应的注册特征。

[0042] 进一步的,上述文本半相关的声纹识别系统中所述改良模块中的 $C_i$ 为所述一段验证数字串的第 $i$ 个数字的特征对应的权值;误识率为 $f$ 的情况下,所述一段验证数字串的第 $i$ 个数字单独用于声纹识别的识别率 $r$ 作为所述一段验证数字串的第 $i$ 个数字的验证特征对应的权值,所述一段验证数字串的第 $i$ 个数字为数字“X”,改良模块中的获取识别率 $r$ 包括:

[0043] 获取 $m$ 个人的数字“X”的语音段,每个人至少包含两段语音,分别提取各语音段的特征向量;

[0044] 构建一定数量的正负对,两个特征为同一人即正对,两个特征为不同人为负对;

[0045] 正对是指注册过程输出的特征与验证过程输出的特征一致,负对是指注册过程输出的特征与验证过程输出的特征不一致,特征是从语音中提取的,如果配对的两段语音不

是同一人的即为负对,为同一人的即为正对;

[0046] 分别对一定数量的正负对的特征向量进行打分,其中,正对的特征向量打分大于 $t$ 的数量为 $Tg$ ,正对的特征向量打分小于 $t$ 的数量为 $Tl$ ;负对的特征向量打分大于 $t$ 的数量为 $Fg$ ,负对的特征向量小于 $t$ 的数量为 $Fl$ ,计算误识率为 $f = Fl / (Fg + Fl)$ ;

[0047] 所述打分是注册特征向量、验证特征向量进行相似度进行计算;

[0048] 识别率为 $r = Tl / (Tg + Tl)$ ,其中,所述一段验证数字串的第 $i$ 个数字为数字“X”的 $C_i$ 值即为确定 $f$ 值时的识别率 $r$ 值。

[0049] 进一步的,上述文本半相关的声纹识别系统中所述改良模块中的所述构建一定数量的正负对,其中一定数量为不低10000;误识率的 $f$ 值可根据场景需求行设定。

[0050] 与现有技术相比,通过本发明的文本半相关的声纹识别方法及系统,即提高声纹识别精度,又可以从较短语音进行声纹识别,同时可以防止录音等情况产生,确保声纹识别应用安全的。

## 附图说明

[0051] 此处所说明的附图用来提供对发明的进一步理解,构成本发明的一部分,本发明的示意性实施例及其说明用于解释本发明,并不构成对本发明的不当限定,显而易见地,下面描述中的附图是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。在附图中:

[0052] 图1为本发明的实施例1文本半相关的声纹识别方法的流程步骤图;

[0053] 图2为本发明的实施例3文本半相关的声纹识别系统的原理示意图。

## 具体实施方式

[0054] 为了使本发明所要解决的技术问题、技术方案及有益效果更加清楚、明白,以下结合附图和实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0055] 本发明实施例提供了一种文本半相关的声纹识别方法,如附图1所示,所述方法包括如下:

[0056] 步骤S1,在注册过程中获取注册用户读几段不同的字符串的语音,对所述语音根据内容进行切分,并对切分后语音段进行特征提取;所述切分是将语音内容的每个字符切成一语音段;

[0057] 步骤S2,随机产生一段验证字符串,所述一段验证字符串包含 $n$ 个字符,

[0058] 步骤S3,获取待验证用户读的一段验证字符串语音内容,并对一段验证字符串语音内容进行切分,分别对每个字符提取的特征,并将提取的特征依次串起来获取验证特征向量 $V$ ;

[0059] 步骤S4,对验证特征向量 $V$ 进行加权获取改良验证特征向量 $V'$ ,其中, $V = [v_1, v_2, \dots, v_n]$ , $V' = [C_1v_1, C_2v_2, \dots, C_nv_n]$ , $v_i$ 为所述一段验证字符串的第 $i$ 个字符的特征, $C_i$ 为所述一段验证字符串的第 $i$ 个字符的特征对应的权值,即 $v_1$ 为所述一段验证字符串的第1个字符的特征, $v_i$ 为所述一段验证字符串的第 $i$ 个字符的特征, $v_n$ 为所述一段验证字符

串的第 $n$ 个字符的特征; $C_1$ 为所述一段验证字符串的第1个字符的特征对应的权值, $C_i$ 为所述一段验证字符串的第 $i$ 个字符串的特征对应的权值, $C_n$ 为所述一段验证字符串的第 $n$ 个字符的特征对应的权值,

[0060] 其中 $i$ 、 $n$ 均为正整数, $n \geq 2, 1 \leq i \leq n$ ;

[0061] 进一步的, $C_i$ 为误识率为 $f$ 的情况下,第 $i$ 个数字单独用于声纹识别的识别率 $r$ 作为第 $i$ 个数字的验证特征对应的权值。

[0062] 识别率 $r$ 的获取方法如下:

[0063] 本发明实施例以字符 $Y$ 为例,介绍字符 $Y$ 验证特征对应的权值的 $C$ 的获取方法,具体包括如下:

[0064] (1) 获取 $m$ 个人的字符 $Y$ 的语音段,每个人至少包含两段语音,分别提取各语音段的特征向量。

[0065] (2) 构建一定数量的正负对,两个特征为同一人即正对,两个特征为不同人为负对;为了使得结果越稳定可靠,其中一定数量为不低10000;

[0066] 正对是指注册过程输出的特征与验证过程输出的特征一致,

[0067] 负对是指注册过程输出的特征与验证过程输出的特征不一致,

[0068] 特征是从语音中提取的,如果配对的两段语音不是同一人的即为负对,为同一人的即为正对,

[0069] (3) 分别对一定数量的正负对的特征向量进行打分,其中,正对的特征向量打分大于 $t$ 的数量为 $T_g$ ,正对的特征向量打分小于 $t$ 的数量为 $T_l$ ;负对的特征向量打分大于 $t$ 的数量为 $F_g$ ,负对的特征向量小于 $t$ 的数量为 $F_l$ ,计算误识率为 $f = F_l / (F_g + F_l)$ ;

[0070] 所述打分是注册特征向量、验证特征向量进行相似度进行计算,其中相似度计算在本发明实施例优选地为,余弦距离 $\cosine$ 计算, $llr = \cos(\text{注册特征向量}, \text{验证特征向量})$ ,不限于这个方法,也可以使用 $plda$ ,欧式距离等多种打分方法。

[0071] (4) 识别率为 $r = T_l / (T_g + T_l)$ ,其中,字符 $Y$ 的 $C$ 值即为确定 $f$ 值时的识别率 $r$ 值,误识率的 $f$ 值可根据场景需求行设定。

[0072] 步骤S5,根据所述一段验证数字串对应的数字获取注册用户在注册过程中对应的特征,并依次即将所述一段验证数字串对应的数字特征串起来获得注册特征向量 $U$ , $U = [u_1, u_2, \dots, u_n]$ ,其中, $u_i$ 为所述一段验证数字串的第 $i$ 个数字对应的注册特征, $u_1$ 为所述一段验证数字串的第1个数字对应的注册特征, $u_2$ 为所述一段验证数字串的第2个数字对应的注册特征, $u_n$ 为所述一段验证数字串的第 $n$ 个数字对应的注册特征;

[0073] 步骤S6,计算改良后的验证特征向量 $V'$ 和注册特征向量 $U$ 的相似度 $llr$ ,即进行余弦距离 $\cosine$ 计算, $llr = \cos(U, V')$ ;

[0074] 步骤S7,根据所述相似度 $llr$ 值的大小判断验证用户与注册用户是否一致。

[0075] 实施例2

[0076] 本发明实施例提供了一种文本半相关的声纹识别方法,如附图2所示,所述方法包括如下:

[0077] 步骤一,在注册过程中获取注册用户读几段不同的字符串的语音,对所述语音根据内容进行切分,并对切分后语音段进行特征提取;所述切分是将语音内容的每个数字切成一语音段;在本发明实施例中,优选地,所述字符串为数字串;需要说明的是,本发明字

字符串不限于数字串,可以为字符串可以为数字、字母、特殊符号、汉字等。

[0078] 步骤二,随机产生一段验证数字串,所述一段验证数字串包含 $n$ 个数字,并获取待验证用户读的一段验证数字串语音内容,并对一段验证数字串语音内容进行切分,分别对每个数字提取的特征,并将提取的特征依次串起来获取验证特征向量 $V$ ,对验证特征向量 $V$ 进行加权获取改良验证特征向量 $V'$ ,其中, $V=[v_1, v_2, \dots, v_n]$ , $V'=[C_1v_1, C_2v_2, \dots, C_nv_n]$ , $i, n$ 均为正整数, $n \geq 2, 1 \leq i \leq n$ ;

[0079] 其中 $v_i$ 为所述一段验证数字串的第 $i$ 个数字的特征, $C_i$ 为所述一段验证数字串的第 $i$ 个数字的特征对应的权值,即 $v_1$ 为所述一段验证数字串的第1个数字的特征, $v_2$ 为所述一段验证数字串的第2个数字的特征,……, $v_n$ 为所述一段验证数字串的第 $n$ 个数字的特征, $C_1$ 为所述一段验证数字串的第1个数字的特征对应的权值, $C_2$ 为所述一段验证数字串的第2个数字的特征对应的权值,……, $C_n$ 为所述一段验证数字串的第 $n$ 个数字的特征对应的权值;

[0080] 其中 $i, n$ 均为正整数, $n \geq 2, 1 \leq i \leq n$ ;

[0081] 进一步的, $C_i$ 为误识率为 $f$ 的情况下,所述一段验证数字串的第 $i$ 个数字单独用于声纹识别的识别率 $r$ 作为所述一段验证数字串的第 $i$ 个数字的验证特征对应的权值。

[0082] 识别率 $r$ 的获取方法如下:

[0083] 本发明实施例以数字‘0’为例,介绍数字0验证特征对应的权值的 $C$ 的获取方法,具体包括如下:

[0084] (1) 获取 $m$ 个人的数字‘0’的语音段,每个人至少包含两段语音,分别提取各语音段的特征向量。

[0085] (2) 构建一定数量的正负对,两个特征为同一人即正对,两个特征为不同人为负对;为了使得结果越稳定可靠,其中一定数量为不低10000;

[0086] 正对是指注册过程输出的特征与验证过程输出的特征一致,

[0087] 负对是指注册过程输出的特征与验证过程输出的特征不一致,

[0088] 特征是从语音中提取的,如果配对的两段语音不是同一人的即为负对,为同一人的即为正对

[0089] (3) 分别对一定数量的正负对的特征向量进行打分,其中,正对的特征向量打分大于 $t$ 的数量为 $T_g$ ,正对的特征向量打分小于 $t$ 的数量为 $T_l$ ;负对的特征向量打分大于 $t$ 的数量为 $F_g$ ,负对的特征向量小于 $t$ 的数量为 $F_l$ ,计算误识率为 $f = F_l / (F_g + F_l)$ ;

[0090] 所述打分是注册特征向量、验证特征向量进行相似度进行计算,其中相似度计算在本发明实施例优选地为,余弦距离 $\cos$ 计算, $11r = \cos(\text{注册特征向量}, \text{特征2})$ ,不限于这个方法,也可以使用 $plda$ ,欧式距离等多种打分方法

[0091] (4) 识别率为 $r = T_l / (T_g + T_l)$ ,其中,数字‘0’的 $C$ 值即为确定 $f$ 值时的识别率 $r$ 值,误识率的 $f$ 值可根据场景需求行设定。另外,识率 $f$ 值由整体声纹识别的需求决定,即误识率的 $f$ 值可根据场景需求行设定。

[0092] 步骤三,根据所述一段验证数字串对应的数字获取注册用户在注册过程中对应的特征,并依次即将所述一段验证数字串对应的数字特征串起来获得注册特征向量 $U$ , $U=[u_1, u_2, \dots, u_n]$ ,其中, $u_i$ 为所述一段验证数字串的第 $i$ 个数字对应的注册特征, $u_1$ 为所述一段验证数字串的第1个数字对应的注册特征, $u_2$ 为所述一段验证数字串的第2个数字对应的

注册特征,  $un$  为所述一段验证数字串的第  $n$  个数字对应的注册特征;

[0093] 步骤四, 计算改良后的验证特征向量  $V'$  和注册特征向量  $U$  的相似度  $llr$ , 即进行进行余弦距离  $\cosine$  计算,  $llr = \cos(U, V')$ ;

[0094] 步骤五, 根据所述相似度  $llr$  值的大小判断验证用户与注册用户是否一致。

[0095] 需要进一步说明的是, 数字  $i$  对应的注册特征是指在注册用户注册过程中获取注册用户的语音内容中数字  $i$  提取的特征; 数字  $i$  对应的验证特征是指在待验证用户在验证过程中获取待验证用户语音内容中数字  $i$  提取的特征。

[0096] 实施例3

[0097] 另外, 本发明实施例还提供了一种文本半相关的声纹识别系统, 如附图2所示, 所述系统包括注册模块1、验证模块2、验证特征模块4、改良模块5、注册特征模块3、判断模块6。

[0098] 所述注册模块用于提供几段不同的字符串供注册用户读取, 获取注册用户读几段不同的字符串的语音, 用于对语音进行切分后语音段进行特征提取; 所述切分是将语音内容的每个字符切成一语音段;

[0099] 所述验证模块2用于随机产生一段验证字符串, 所述一段验证字符串包含  $n$  个字符;

[0100] 所述验证特征模块4, 用于获取待验证用户读的一段验证字符串语音内容, 并对一段验证字符串语音内容进行切分, 分别对每个字符提取的特征, 并将提取的特征依次串起来获取验证特征向量  $V, V = [v_1, v_2, \dots, v_n]$ ,  $v_i$  为所述一段验证字符串的第  $i$  个字符的特征,  $v_1$  为所述一段验证字符串的第1个字符的特征,  $v_2$  为所述一段验证字符串的第2个字符的特征,  $v_n$  为所述一段验证字符串的第  $n$  个字符的特征, 其中  $i, n$  均为正整数,  $n \geq 2, 1 \leq i \leq n$ ;

[0101] 所述改良模块5, 用于对验证特征向量  $V$  进行加权获取改良验证特征向量  $V', V' = [C_1v_1, C_2v_2, \dots, C_nv_n]$ ,  $C_i$  为所述一段验证字符串的第  $i$  个字符的特征对应的权值,  $C_1$  为所述一段验证字符串的第1个字符的特征对应的权值,  $C_2$  为所述一段验证字符串的第2个字符的特征对应的权值,  $C_n$  为所述一段验证字符串的第  $n$  个字符的特征对应的权值。

[0102] 所述注册特征模块3用于根据所述验证模块2产生的所述一段验证字符串对应的字符获取注册用户在注册过程中对应的特征, 并依次即将所述一段验证字符串对应的特征串起来获得注册特征向量  $U, U = [u_1, u_2, \dots, u_n]$ , 其中,  $u_i$  为所述一段验证字符串的第  $i$  个字符对应的注册特征,  $u_1$  为所述一段验证字符串的第1个字符对应的注册特征,  $u_2$  为所述一段验证字符串的第2个字符对应的注册特征,  $u_n$  为所述一段验证字符串的第  $n$  个字符对应的注册特征;

[0103] 所述判断模块6用于计算改良后的验证特征向量  $V'$  和注册特征向量  $U$  的相似度  $llr$ , 并根据所述相似度  $llr$  值的大小判断验证用户与注册用户是否一致。

[0104] 在本发明实施例中, 优选地, 所述字符串为数字串;

[0105] 所述注册模块1提供几段不同的数字串供注册用户读取, 所述几段不同的数字串的语音为几段不同的数字串的语音, 所述切分是将语音内容的每个数字切成一语音段;

[0106] 所述验证模块2随机产生一段验证数字串, 所述一段验证数字串包含  $n$  个数字;

[0107] 所述验证特征模块4获取待验证用户读的一段验证数字串语音内容, 并对一段验

证数字串语音内容进行切分,分别对每个数字提取的特征,并将提取的特征依次串起来获取验证特征向量 $V$ , $V=[v_1,v_2,\dots,v_n]$ , $v_i$ 为所述一段验证数字串的第 $i$ 个数字的特征;

[0108] 改良模块5,用于对验证特征向量 $V$ 进行加权获取改良验证特征向量 $V'$ , $V'=[C_1v_1,C_2v_2,\dots,C_nv_n]$ , $C_i$ 为所述一段验证数字串的第 $i$ 个数字的特征对应的权值;

[0109] 注册特征模块,用于根据所述验证模块产生的所述一段验证数字串对应的数字获取注册用户在注册过程中对应的特征,并依次即将所述一段验证数字串对应的特征串起来获得注册特征向量 $U$ , $U=[u_1,u_2,\dots,u_n]$ ,其中, $u_i$ 为所述一段验证数字串的第 $i$ 个数字对应的注册特征, $u_1$ 为所述一段验证数字串的第1个数字对应的注册特征, $u_2$ 为所述一段验证数字串的第2个数字对应的注册特征, $u_n$ 为所述一段验证数字串的第 $n$ 个数字对应的注册特征。

[0110] 改良模块5中的 $C_i$ 为所述一段验证数字串的第 $i$ 个数字的特征对应的权值;误识率为 $f$ 的情况下,所述一段验证数字串的第 $i$ 个数字单独用于声纹识别的识别率 $r$ 作为所述一段验证数字串的第 $i$ 个数字的验证特征对应的权值,所述一段验证数字串的第 $i$ 个数字为数字“X”,改良模块中的获取识别率 $r$ 包括:

[0111] 获取 $m$ 个人的数字“X”的语音段,每个人至少包含两段语音,分别提取各语音段的特征向量;

[0112] 构建一定数量的正负对,两个特征为同一人即正对,两个特征为不同人为负对;

[0113] 正对是指注册过程输出的特征与验证过程输出的特征一致,负对是指注册过程输出的特征与验证过程输出的特征不一致,特征是从语音中提取的,如果配对的两段语音不是同一人的即为负对,为同一人的即为正对;

[0114] 分别对一定数量的正负对的特征向量进行打分,其中,正对的特征向量打分大于 $t$ 的数量为 $T_g$ ,正对的特征向量打分小于 $t$ 的数量为 $T_l$ ;负对的特征向量打分大于 $t$ 的数量为 $F_g$ ,负对的特征向量小于 $t$ 的数量为 $F_l$ ,计算误识率为 $f=F_l/(F_g+F_l)$ ;

[0115] 所述打分是注册特征向量、验证特征向量进行相似度进行计算,当前相似值算法以 $\cosine$ 为例,则相似值越小两者越可能是同一人,反之则越不是一个人, $t$ 值实际应用中可根据精度需求设置阈值。

[0116] 识别率为 $r=T_l/(T_g+T_l)$ ,其中,所述一段验证数字串的第 $i$ 个数字为数字“X”的 $C_i$ 值即为确定 $f$ 值时的识别率 $r$ 值。

[0117] 进一步的,构建一定数量的正负对,总数量越多结果越稳定可靠,其中一定数量为不低10000。

[0118] 另外,误识率 $f$ 值由整体声纹识别的需求决定,即误识率的 $f$ 值可根据场景需求行设定。

[0119] 需要说明的是,本发明字符串不限于数字串,以可以为字符串可以为数字、字母、特殊符号、汉字等。

[0120] 以上各实施例仅用以说明本发明的技术方案,但应当理解本发明并非局限于上述实施例,通过本发明的启示,本领域技术人员结合公知或现有技术、知识所进行的修改,或者对其中部分或者全部技术特征进行等同替换也应视为在本发明的保护范围内。

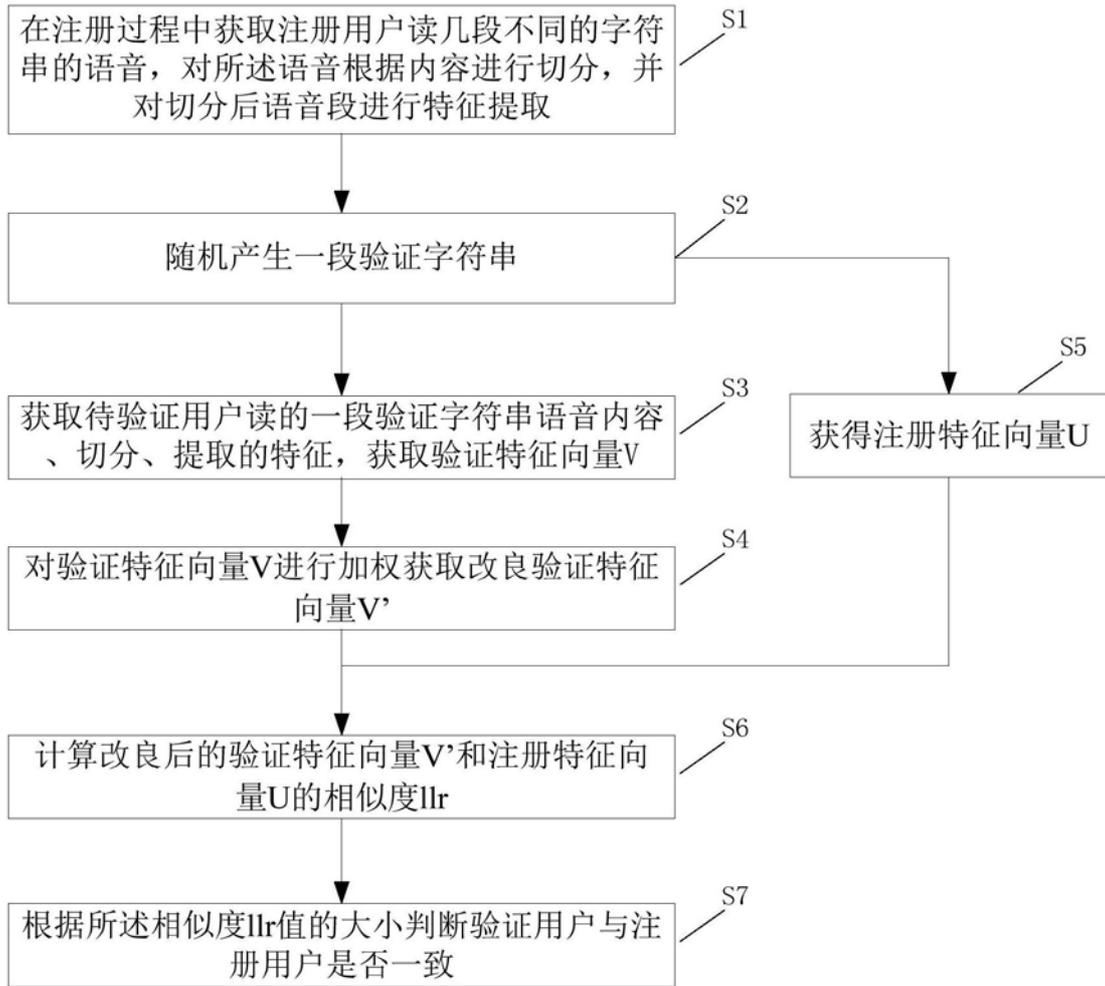


图1

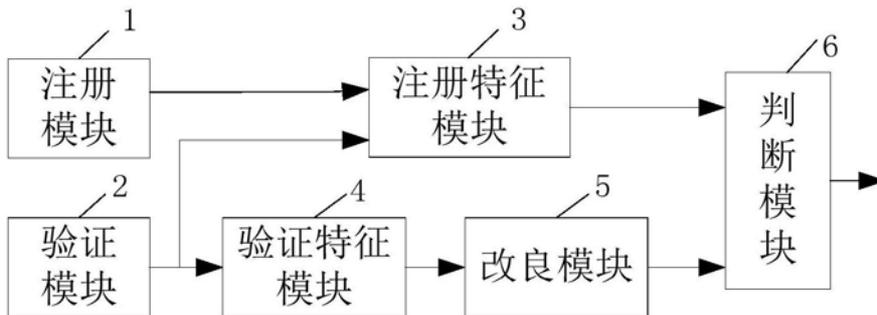


图2