



(12) 发明专利申请

(10) 申请公布号 CN 114357417 A

(43) 申请公布日 2022. 04. 15

(21) 申请号 202111677950.4

(22) 申请日 2021.12.31

(71) 申请人 中国科学院声学研究所东海研究站  
地址 200032 上海市徐汇区小木桥路456号

(72) 发明人 洪峰

(74) 专利代理机构 上海汉声知识产权代理有限公司 31236

代理人 胡晶

(51) Int. Cl.

G06F 21/32 (2013.01)

G06F 40/205 (2020.01)

G06F 16/903 (2019.01)

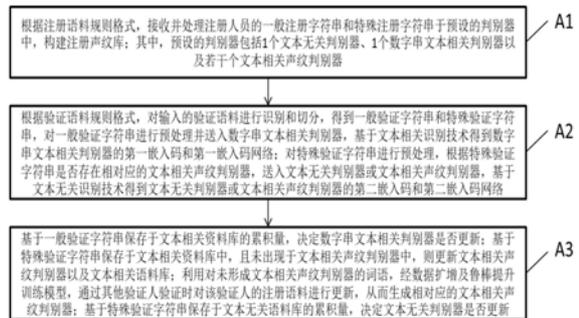
权利要求书4页 说明书9页 附图3页

(54) 发明名称

一种基于未知语料的自学习动态声纹身份验证方法

(57) 摘要

本发明公开了一种基于未知语料的自学习动态声纹身份验证方法,根据注册语料规则格式,接收并处理注册人员的一般注册字符串和少量特殊注册字符串于训练好的各判别器中,构建注册声纹库。根据验证语料规则格式,进行预处理,经前置条件通过后,数字串会基于文本相关识别技术得到第一嵌入码;特殊字串会基于文本无关识别技术得到第二嵌入码,从而查看是否匹配。根据保存于文本相关资料库的累积量、性能提升度、录入语料人数量等因素对文本无关判别器、一般字符串数字串文本相关判别器以及特殊字符串文本相关声纹判别器进行更新。本发明具有较好的性能,基于未知语料提升安全性,基于模型和数据库动态增加自动提升性能。



1. 一种基于未知语料的自学习动态声纹身份验证方法,配置有动态声纹身份验证系统,其特征在于,所述方法包括:

A1:根据注册语料规则格式,接收并处理注册人员的一般注册字符串和所述特殊注册字符串于预设的判别器中,构建注册声纹库;其中,预设的判别器包括1个文本无关判别器、1个一般字符串数字串文本相关判别器以及若干个特殊字符串文本相关声纹判别器;

A2:根据验证语料规则格式,对输入的验证语料进行识别和二元切分,得到一般验证字符串和特殊验证字符串,并经过各自的语音识别引擎识别比对,如通过后,再对所述一般验证字符串进行预处理并送入所述一般字符串数字串文本相关判别器,基于文本相关识别技术得到所述一般字符串数字串文本相关判别器的第一嵌入码;对所述特殊验证字符串进行预处理,根据特殊验证字符串是否存在相对应的所述特殊字符串文本相关声纹判别器以及是否存在至少一条注册嵌入码,送入所述文本无关判别器或所述特殊字符串文本相关声纹判别器,基于文本无关识别技术得到所述文本无关判别器或所述文本相关声纹判别器的第二嵌入码;

A3:根据保存于文本相关资料库的所述一般验证字符串的累积量以及性能提升度同时满足一定阈值,则更新所述一般字符串数字串文本相关判别器;

根据所述特殊验证字符串保存于所述文本相关资料库中的录入语料人数量及其语料数量同时满足一定阈值,则更新所述文本相关语料库以及所述特殊字符串文本相关声纹判别器;

根据所述特殊验证字符串保存于文本无关语料库的累积量以及性能提升度同时满足一定阈值,则更新所述文本无关判别器。

2. 根据权利要求1所述的基于未知语料的自学习动态声纹身份验证方法,其特征在于,在所述步骤A2验证过程中,还会利用未形成所述特殊字符串文本相关声纹判别器的词语,根据数据积累量以及性能提升度满足一定阈值后,经数据扩增及鲁棒处理后,具备未形成所述特殊字符串文本相关声纹判别器训练模型所需的文本相关语料库,其获取方式为通过部分验证人验证时对所述文本相关语料库进行更新增加,使得录入语料人数量及其语料数量同时满足一定阈值,从而经训练及性能验证生成相对应新的所述特殊字符串文本相关声纹判别器。

3. 根据权利要求1所述的基于未知语料的自学习动态声纹身份验证方法,其特征在于,所述步骤A1包括:

S11:输入所述一般注册字符串和所述特殊注册字符串至所述动态声纹身份验证系统,所述一般注册字符串和所述特殊注册字符串分别经所述一般字符串数字串文本相关判别器和所述特殊字符串文本相关声纹判别器所形成的前端音频特征实现判别处理,而后存储于所述文本相关语料库中;

S12:输入验证语料,对所述验证语料进行二元切分,得到所述一般验证字符串和所述特殊验证字符串,通过文本相关识别技术得到第一嵌入码,所述第一嵌入码可由所述一般字符串数字串文本相关判别器提取,基于文本无关识别技术得到第二嵌入码,所述第二嵌入码可由所述特殊字符串文本相关声纹判别器提取。

4. 根据权利要求1所述的基于未知语料的自学习动态声纹身份验证方法,其特征在于,所述步骤A2包括:

S2: 根据所述动态声纹身份验证系统的动态验证规则, 输入所述动态声纹身份验证系统给出的一动态验证语料, 对所述动态验证语料进行二元切分, 得到一般动态验证字符串和特殊动态验证字符串, 并分别经过各自的语音识别引擎进行识别比对, 若通过后, 判断为所述一般动态验证字符串依次进行语速测试、重新排列以及补齐重建等预处理手段, 并进入步骤S3进行后续处理, 判断为所述特殊动态验证字符串则先经过语音识别引擎判断是否同音, 若是则再进入步骤S4进入后续处理, 其中, 所述一般动态验证字符串由若干数字组成, 所述特殊动态验证字符串为中文词语或英文字母或英文词组组成;

S3: 将所述数字串送入预设的一般字符串数字串文本相关判别器中与预设阈值进行判别, 若判别成功, 将所述数字串同时送入文本无关语料库和文本相关资料库进行保存; 否则认定为验证失败并停止;

S4: 判断所述特殊动态验证字符串是否存在相对应的所述特殊字符串文本相关声纹判别器, 以及所述文本相关语料库是否存在待验证者所述特殊动态验证字符串的至少1条注册语料, 若两者均存在则进入步骤S5, 若相对应的所述特殊字符串文本相关声纹判别器不存在或两者均不存在则进入步骤S6, 若相对应的所述特殊字符串文本相关声纹判别器存在但未激活以及所述文本相关语料库内至少1条注册语料存在, 则进入步骤S7;

S5: 将所述特殊动态验证字符串送入相对应的所述特殊字符串文本相关声纹判别器与待验证者的注册嵌入码进行相似性判别, 若相似性超过预设阈值, 则判定成功, 将所述特殊动态验证字符串送入所述文本无关语料库和所述文本相关资料库进行保存, 否则认定为验证失败并停止;

S6: 将所述特殊动态验证字符串送入所述文本无关判别器与待验证者的注册嵌入码进行相似性判别, 若相似性超过预设阈值, 若判定成功, 则送入所述文本无关语料库和文本相关语料库进行保存, 否则认定为验证失败并停止;

S7: 将所述特殊动态验证字符串送入所述文本无关判别器进行相似性判别, 若相似性超过预设阈值, 若判定成功, 则送入所述文本无关语料库进行保存, 并对所述文本相关语料库进行后处理, 相对应的所述特殊字符串文本相关声纹判别器被激活, 新增至少一条可被所述特殊字符串文本相关声纹判别器判别的注册语料, 否则认定为验证失败并停止。

5. 根据权利要求4所述的基于未知语料的自学习动态声纹身份验证方法, 其特征在于, 在所述步骤S2中, 所述对所述动态验证语料进行切分之前还包括如下步骤:

B1: 对所述动态验证语料进行环境语音质量检测判断环境噪音水平, 若计算得到的信噪比与预设阈值进行比较, 若低于预设阈值则进入步骤B2, 否则进入步骤B3;

B2: 提示再次输入所述动态验证语料, 并返回至所述步骤B1, 进入所述步骤B2次数为 $N$   
 $=N+1, N=0$ , 若 $N>3$ 则认定为验证失败并停止, 其中,  $N>3$ 的判定条件为可调整的;

B3: 将 $N$ 清零, 并将所述动态验证语料转变为不局限于wav格式从而进行切分;

进一步地, 在所述步骤B1中能为进行环境语音质量检测前的所述动态验证语料进行视觉上加噪处理。

6. 根据权利要求4所述的基于未知语料的自学习动态声纹身份验证方法, 其特征在于, 所述步骤S2具体包括如下步骤:

S21: 对所述动态验证字符串语音通过语音识别引擎进行数字识别, 并通过语速检测器进行判断, 若识别通过且语速正常则进入步骤S22, 否则认定为验证失败并停止;

S22:对所述动态验证字符串进行完全切分,得到若干仅包含一个数字的数字段;

S23:对若干所述数字段基于所述一般注册字符串的排列顺序进行重新排列;

S24:根据先验规则调整补齐缺失语料,对排列后的所述数字段进行补齐重建得到所述数字串。

7.根据权利要求4所述的基于未知语料的自学习动态声纹身份验证方法,其特征在于,所述步骤S3具体包括如下步骤:

S31:将所述数字串送入所述一般字符串数字串文本相关判别器中进行判别,得到第一验证嵌入码,并将所述第一验证嵌入码与所述第一嵌入码进行求解以获取所述数字串相似性,以进行声纹匹配,若所述数字串相似性超过预设阈值,则判定成功进入步骤S32,否则认定为验证失败并停止;

S32:将所述数字串送入所述文本相关语料库和所述文本无关语料库进行保存,将所述第一验证嵌入码和其余所述第一嵌入码经计算后形成新的第一嵌入码。

8.根据权利要求4所述的基于未知语料的自学习动态声纹身份验证方法,其特征在于,所述步骤S5具体包括如下步骤:

S51:将所述特殊动态验证字符串送入相对应的所述特殊字符串文本相关声纹判别器生成得到第二验证嵌入码,并将所述第二验证嵌入码与所述第二嵌入码进行求解以获取所述特殊动态验证字符串的相似性,以进行声纹匹配,若所述特殊动态验证字符串的相似性超过预设阈值,则判定成功进入步骤S52,否则认定为验证失败并停止;

S52:将所述特殊动态验证字符串送入所述文本无关语料库保存和所述文本相关语料库进行保存,将所述第二验证嵌入码和其余所述第二嵌入码经计算后形成新的第二嵌入码。

9.根据权利要求4所述的基于未知语料的自学习动态声纹身份验证方法,其特征在于,所述步骤S6具体包括如下步骤:

S61:将所述特殊动态验证字符串送入所述文本无关判别器,生成第二验证嵌入码,并将所述第二验证嵌入码与所述第二嵌入码,可以语句级或帧级层面上,进行求解以获取所述特殊动态验证字符串的相似性;

S62:对所述特殊动态验证字符串进行语音识别引擎判定,若满足阈值则进入所述步骤S63,否则认定为验证失败并停止;若验证者注册语料存在则直接进入所述步骤S63;

S63:对所述特殊动态验证字符串进行声纹匹配,若匹配成功则进入步骤S64,否则认定为验证失败并停止;

S64:将所述特殊动态验证字符串送入所述文本无关语料库保存,生成特殊字符串相关语料送入临时语料库中,当所述特殊字符串相关语料在所述临时语料库中长度以及录入该试验语料人数均超过预设阈值时,可形成特殊字符串相关语料库,用于训练新的相关判别器;

S65:将所述特殊字符串相关语料送入所述文本相关语料库进行保存,并对所述特殊字符串相关语料进行训练形成新的特殊字符串文本相关声纹判别器。

10.根据权利要求4所述的基于未知语料的自学习动态声纹身份验证方法,其特征在于,所述步骤S7具体包括如下步骤:

S71:将所述特殊动态验证字符串送入所述文本无关判别器,生成第二验证嵌入码,并

将所述第二验证嵌入码与所述第二嵌入码进行求解以获取所述特殊动态验证字符串的相似性,以进行声纹匹配,若满足预设阈值则进入所述步骤S72,否则认定为验证失败并停止;

S72:将所述特殊动态验证字符串送入所述文本无关语料库保存,并对所述文本相关语料库实现后处理;

S73:对所述文本相关语料库中相对应的特殊字符串文本相关声纹判别器进行激活,新增至少一条可被所述特殊字符串文本相关声纹判别器判别的注册语料。

11.根据权利要求1所述的基于未知语料的自学习动态声纹身份验证方法,其特征在于,在所述步骤A3还包括:

对于所述文本相关语料库,以特定验证人为数据组织方式,将其所有语料以对抗训练方式训练相关的声纹合成网络;以既有语料文本作为输入,形成音频作为扩增的语料文件,并添加噪声信号,形成所述文本相关语料库内新的语料文件。

## 一种基于未知语料的自学习动态声纹身份验证方法

### 技术领域

[0001] 本发明属于声纹技术的身份验证领域,尤其涉及一种基于未知语料的自学习动态声纹身份验证方法。

### 背景技术

[0002] 声纹识别是利用人的声纹特征来进行身份认证的生物特征识别技术,具有采集成本低、实现简单、体验性好等优点。更为重要的是,声纹特征是所有生物特征中唯一的低敏感度信息,与人脸识别、指纹识别等其他生物特征识别技术相比,声纹识别涉及的个人隐私信息少,更容易被用户接受。按识别内容可将声纹识别分为文本无关、文本相关和文本提示型三种。然而目前基于这三种识别方案的软件均缺乏反欺诈功能,使得冒充者有可能通过各种手段伪造出和目标说话人接近的语音,从而对声纹识别系统进行欺诈攻击,例如在微信和支付宝推出的声音锁中均存在此类安全隐患。对于文本相关的声纹识别,常用的固定式验证码容易被录音欺诈,即冒充者播放已经录制好的目标用户的声音通过声纹验证。文本提示型的声纹识别虽然具有一定的反欺诈功能,但其受文本集合的限制较大,文本集合太小,则识别系统的安全性不够;文本集合太大,则用户需要录入大量的音频,体验感降低。此外,文本提示型的声纹识别还缺乏反录音拼接欺诈功能,比如将用户对固定输入0-9的数字或者A-Z字母的发音全部采集起来,切割出目标说话人说单个数字或字母的语音并形成文本语音库,在进行声纹确认时,利用计算机程序可以根据文本语音库迅速地拼接出目标验证码的发音。

[0003] 因此,亟需提出一种技术,能够在不涉及更多的个人隐私以及不降低用户体验的基础之上,提升了声纹识别系统的安全性,以有效解决在各类软件登录或金融支付过程中基于声纹识别的身份认证方案中的安全性差的问题。

### 发明内容

[0004] 本发明的技术目的是提供一种基于未知语料的自学习动态声纹身份验证方法,以解决声纹识别安全性差的技术问题。

[0005] 为解决上述问题,本发明的技术方案为:

[0006] 一种基于未知语料的自学习动态声纹身份验证方法,配置有动态声纹身份验证系统,方法包括:

[0007] A1:根据注册语料规则格式,接收并处理注册人员的一般注册字符串和特殊注册字符串于预设的判别器中,构建注册声纹库;其中,预设的判别器包括1个文本无关判别器、1个一般字符串数字串文本相关判别器以及若干个特殊字符串文本相关声纹判别器;

[0008] A2:根据验证语料规则格式,对输入的验证语料进行识别和二元切分,得到一般验证字符串和特殊验证字符串,并经过各自的语音识别引擎识别比对,如通过后,再对一般验证字符串进行预处理并送入一般字符串数字串文本相关判别器,基于文本相关识别技术得到一般字符串数字串文本相关判别器的第一嵌入码;对特殊验证字符串进行预处理,根据

特殊验证字符串是否存在相对应的特殊字符串文本相关声纹判别器以及是否存在至少一条注册嵌入码,送入文本无关判别器或特殊字符串文本相关声纹判别器,基于文本无关识别技术得到文本无关判别器或文本相关声纹判别器的第二嵌入码;

[0009] A3:根据保存于文本相关资料库的一般验证字符串的累积量以及性能提升度同时满足一定阈值,则更新一般字符串数字串文本相关判别器;

[0010] 根据特殊验证字符串保存于文本相关资料库中的录入语料人数量及其语料数量同时满足一定阈值,则更新文本相关语料库以及特殊字符串文本相关声纹判别器;

[0011] 根据特殊验证字符串保存于文本无关语料库的累积量以及性能提升度同时满足一定阈值,则更新文本无关判别器。

[0012] 进一步优选地,在步骤A2验证过程中,还会利用未形成特殊字符串文本相关声纹判别器的词语,根据数据积累量以及性能提升度满足一定阈值后,经数据扩增及鲁棒处理后,具备未形成特殊字符串文本相关声纹判别器训练模型所需的文本相关语料库,其获取方式为通过部分验证人验证时对文本相关语料库进行更新增加,使得录入语料人数量及其语料数量同时满足一定阈值,从而经训练及性能验证生成相对应新的特殊字符串文本相关声纹判别器。

[0013] 具体地,步骤A1包括:

[0014] S11:输入一般注册字符串和特殊注册字符串至动态声纹身份验证系统,一般注册字符串和特殊注册字符串分别经一般字符串数字串文本相关判别器和特殊字符串文本相关声纹判别器所形成的前端音频特征实现判别处理,而后存储于文本相关语料库中;

[0015] S12:输入验证语料,对验证语料进行二元切分,得到一般验证字符串和特殊验证字符串,通过文本相关识别技术得到第一嵌入码,第一嵌入码可由一般字符串数字串文本相关判别器提取,基于文本无关识别技术得到第二嵌入码,第二嵌入码可由特殊字符串文本相关声纹判别器提取。

[0016] 具体地,步骤A2包括:

[0017] S2:根据动态声纹身份验证系统的动态验证规则,输入动态声纹身份验证系统给出的一动态验证语料,对动态验证语料进行二元切分,得到一般动态验证字符串和特殊动态验证字符串,并分别经过各自的语音识别引擎进行识别比对,若通过后,判断为一般动态验证字符串依次进行语速测试、重新排列以及补齐重建等预处理手段,并进入步骤S3进行后续处理,判断为特殊动态验证字符串则先经过语音识别引擎判断是否同音,若是则再进入步骤S4进入后续处理,其中,一般动态验证字符串由若干数字组成,特殊动态验证字符串为中文词语或英文字母或英文词组组成;

[0018] S3:将数字串送入预设的一般字符串数字串文本相关判别器中与预设阈值进行判别,若判别成功,将数字串同时送入文本无关语料库和文本相关资料库进行保存,否则认定为验证失败并停止;

[0019] S4:判断特殊动态验证字符串是否存在相对应的特殊字符串文本相关声纹判别器,以及文本相关语料库是否存在待验证者特殊动态验证字符串的至少1条注册语料,若两者均存在则进入步骤S5,若相对应的特殊字符串文本相关声纹判别器不存在或两者均不存在则进入步骤S6,若相对应的特殊字符串文本相关声纹判别器存在但未激活以及文本相关语料库内至少1条注册语料存在,则进入步骤S7;

[0020] S5:将特殊动态验证字符串送入相对应的特殊字符串文本相关声纹判别器进行相似性判别,若相似性超过预设阈值,则判定成功,将特殊动态验证字符串送入文本无关语料库和文本相关资料库进行保存,否则认定为验证失败并停止;

[0021] S6:将特殊动态验证字符串送入文本无关判别器进行相似性判别,若相似性超过预设阈值,若判定成功,则送入文本无关语料库和文本相关语料库进行保存,否则认定为验证失败并停止;

[0022] S7:将特殊动态验证字符串送入文本无关判别器进行相似性判别,若相似性超过预设阈值,若判定成功,则送入文本无关语料库进行保存,并对文本相关语料库进行后处理,相对应的特殊字符串文本相关声纹判别器被激活,新增至少一条可被所述特殊字符串文本相关声纹判别器判别的注册语料,新增一条用于提取注册码,否则认定为验证失败并停止。

[0023] 进一步优选地,在步骤S2中,对动态验证语料进行切分之前还包括如下步骤:

[0024] B1:对动态验证语料进行环境语音质量检测判断环境噪音水平,若计算得到的信噪比与预设阈值进行比较,若低于预设阈值则进入步骤B2,否则进入步骤B3;

[0025] B2:提示再次输入动态验证语料,并返回至步骤B1,进入步骤B2次数为 $N=N+1$ , $N=0$ ,若 $N>3$ 则认定为验证失败并停止,其中, $N>3$ 的判定条件为可调整的;

[0026] B3:将N清零,并将动态验证语料转变为不局限于wav格式从而进行切分;

[0027] 进一步地,在步骤B1中能为进行环境语音质量检测前的动态验证语料进行加噪处理。

[0028] 其中,步骤S2具体包括如下步骤:

[0029] S21:对动态验证字符串语音通过语音识别引擎进行数字识别,并通过语速检测器进行判断,若识别通过且语速正常则进入步骤S22,否则认定为验证失败并停止;

[0030] S22:对动态验证字符串进行完全切分,得到若干仅包含一个数字的数字段;

[0031] S23:对若干数字段基于一般注册字符串的排列顺序进行重新排列;

[0032] S24:根据先验规则调整补齐缺失语料,对排列后的数字段进行补齐重建得到数字串。

[0033] 其中,步骤S3具体包括如下步骤:

[0034] S31:将数字串送入一般字符串数字串文本相关判别器中进行判别,生成得到第一验证嵌入码,并将第一验证嵌入码与第一嵌入码进行求解以获取数字串相似性,以进行声纹匹配,若数字串相似性超过预设阈值,则判定成功进入步骤S32,否则认定为验证失败并停止;

[0035] S32:将数字串送入文本相关语料库和文本无关语料库进行保存,将第一验证嵌入码和其余第一嵌入码经计算后形成新的第一嵌入码。

[0036] 其中,步骤S5具体包括如下步骤:

[0037] S51:将特殊动态验证字符串送入相对应的特殊字符串文本相关声纹判别器生成得到第二验证嵌入码,并将第二验证嵌入码与第二嵌入码进行求解以获取特殊动态验证字符串的相似性,以进行声纹匹配,若特殊动态验证字符串的相似性超过预设阈值,则判定成功进入步骤S52,否则认定为验证失败并停止;

[0038] S52:将特殊动态验证字符串送入文本无关语料库和文本相关语料库进行保存,将

第二验证嵌入码和其余第二嵌入码经计算后形成新的第二嵌入码。

[0039] 其中,步骤S6具体包括如下步骤:

[0040] S61:将所述特殊动态验证字符串送入所述文本无关判别器,生成得到第二验证嵌入码,并将所述第二验证嵌入码与所述第二嵌入码,可以语句级或帧级层面上,进行求解以获取所述特殊动态验证字符串的相似性;

[0041] S62:对所述特殊动态验证字符串进行语音识别引擎判定,若满足阈值则进入所述步骤S63,否则认定为验证失败并停止;若验证者注册语料存在则直接进入所述步骤S63;

[0042] S63:对特殊动态验证字符串进行声纹匹配,若匹配成功则进入步骤S64,否则认定为验证失败并停止;

[0043] S64:将特殊动态验证字符串送入文本无关语料库保存,生成特殊字符串相关语料送入临时语料库中,当特殊字符串相关语料在临时语料库中长度以及录入该试验语料人数均超过预设阈值时,可形成特殊字符串相关语料库,用于训练新的相关判别器;

[0044] S65:将特殊字符串相关语料送入文本相关语料库进行保存,并对特殊字符串相关语料进行训练形成新的特殊字符串文本相关声纹判别器。

[0045] 其中,步骤S7具体包括如下步骤:

[0046] S71:将特殊动态验证字符串送入文本无关判别器,生成第二验证嵌入码,并将第二验证嵌入码与第二嵌入码进行求解以获取特殊动态验证字符串的相似性,若满足预设阈值则进入步骤S72,否则认定为验证失败并停止;

[0047] S72:将特殊动态验证字符串送入文本无关语料库保存,并对文本相关语料库;

[0048] S73:对文本相关语料库中相对应的特殊字符串文本相关声纹判别器进行激活,新增至少一条可被特殊字符串文本相关声纹判别器判别的注册语料。

[0049] 进一步优选地,在步骤A3还包括:

[0050] 对于文本相关语料库,以特定验证人为数据组织方式,将其所有语料以对抗训练方式训练相关的声纹合成网络;以既有语料文本作为输入,形成音频作为扩增的语料文件,并添加噪声信号,形成文本相关语料库内新的语料文件。

[0051] 本发明由于采用以上技术方案,使其与现有技术相比具有以下优点和积极效果:

[0052] 该发明结合了语音识别技术和声纹识别技术,具有较低的错误接受率和错误拒绝率。

[0053] 采用新的注册和验证方式,分为一般+特殊(例如,数字+汉字)的验证码方式,添加一定的视觉干扰,结合声纹识别技术,提高安全性并且本发明充分发挥了文本相关和文本无关声纹识别的优势,具有反录音欺诈和反录音拼接欺诈功能。另外本发明基于大量未知语料,很大程度上提升安全性。

[0054] 本发明对输入的验证语料进行视觉上加噪,提高安全性和防窃性。

[0055] 文本无关判别器、一般字符串数字串文本相关判别器以及特殊字符串文本相关声纹判别器共同自主判别;对于每个待验证者经过的技术验证流程可能均不同此外,模型和数据库都在动态动态增长更新的,具有自动更新和自学习功能,能够在用户的使用过程中不断对识别系统进行更新,提升对目标说话人识别能力以及对冒充者的防御能力。

## 附图说明

[0056] 通过阅读下文优选实施方式的详细描述,各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的,而并不认为是对本发明的限制。

[0057] 图1为本发明的一种基于未知语料的自学习动态声纹身份验证方法的流程图;

[0058] 图2为本发明的一种基于未知语料的自学习动态声纹身份验证方法的实际流程图;

[0059] 图3为本发明的注册字符串示意图;

[0060] 图4为本发明的验证字符串示意图;

[0061] 图5为本发明模型语料关系示意图。

## 具体实施方式

[0062] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对照附图说明本发明的具体实施方式。显而易见地,下面描述中的附图仅仅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,并获得其他的实施方式。

[0063] 为使图面简洁,各图中只示意性地表示出了与本发明相关的部分,它们并不代表其作为产品的实际结构。另外,以使图面简洁便于理解,在有些图中具有相同结构或功能的部件,仅示意性地绘示了其中的一个,或仅标出了其中的一个。在本文中,“一个”不仅表示“仅此一个”,也可以表示“多于一个”的情形。

[0064] 以下结合附图和具体实施例对本发明提出的一种基于未知语料的自学习动态声纹身份验证方法作进一步详细说明。根据下面说明和权利要求书,本发明的优点和特征将更清楚。

[0065] 实施例

[0066] 参看图1和图2,本实施例提供一种基于未知语料的自学习动态声纹身份验证方法,配置于动态声纹身份验证系统中进行使用,能够通过声音起到解锁、开启的作用。

[0067] 该方法如下A1:根据注册语料规则格式,接收并处理注册人员的一般注册字符串和特殊注册字符串于预设的判别器中,构建注册声纹库;其中,预设的判别器包括1个文本无关判别器、1个一般字符串数字串文本相关判别器以及若干个特殊字符串文本相关声纹判别器。

[0068] A2:根据验证语料规则格式,对输入的验证语料进行识别和二元切分,得到一般验证字符串和特殊验证字符串,并经过各自的语音识别引擎识别比对,如通过后,再对一般验证字符串进行预处理并送入一般字符串数字串文本相关判别器,基于文本相关识别技术得到一般字符串数字串文本相关判别器的第一嵌入码;对特殊验证字符串进行预处理,根据特殊验证字符串是否存在相对应的特殊字符串文本相关声纹判别器以及是否存在至少一条注册嵌入码,送入文本无关判别器或特殊字符串文本相关声纹判别器,基于文本无关识别技术得到文本无关判别器或文本相关声纹判别器的第二嵌入码。

[0069] A3:根据保存于文本相关资料库的一般验证字符串的累积量以及性能提升度同时满足一定阈值,则更新一般字符串数字串文本相关判别器;

[0070] 根据特殊验证字符串保存于文本相关资料库中的录入语料人数量及其语料数量同时满足一定阈值,则更新文本相关语料库以及特殊字符串文本相关声纹判别器。

[0071] 根据特殊验证字符串保存于文本无关语料库的累积量以及性能提升度同时满足一定阈值,则更新文本无关判别器。

[0072] 现根据实际应用对本实施例进行说明

[0073] 首先,参看图3和图4,在进行身份验证前需要先录入相关信息,即注册阶段。

[0074] 在步骤S11、S12中,使用者需要向该系统中先输入注册字符串,注册字符串为语料信息,注册字符串包括一般注册字符串和特殊注册字符串。具体地,使用者注册时仅需K组注册字符串(一般 $K \leq 5$ ),每组注册字符串的内容随机不同、但字符个数相同。前T组为一般注册字符串,则K-T组为特殊字符串。通常地,一般字符串由数字字符构成,特殊字符串由字母或英文词组或汉字字符构成,以此构建注册声纹库,每个人都在该注册声纹库具有一个独立的部分。

[0075] 在注册前,已训练有同时满足FRR和FAR要求的若干组判别器,包括1个一般字符串数字串文本相关判别器、若干个特殊的特殊字符串文本相关声纹判别器(在本实施例可为100个),1个文本无关判别器。

[0076] 参看图3为方便演示所展示的一组注册字符串,不以其具体内容进行限定。这里设计需要录入 $K=5$ 组语料的注册字符串用于提取初始注册字符,包括一般注册字符串和特殊注册字符串。由字母或英文词组或汉字字符每个字符内容至少重复2遍、每个字符串的字符数均为10个,前2组语料为一般字符串,后3组语料为特殊字符串,其中第3组为字母,后5个字符是前5个字符的重复;第4-5组为中文词组,每个词组均为2个字。

[0077] 然后在步骤S2中,根据身份验证系统给定的验证语料进行输入,系统会对验证语料进行切分为两段,得到一般验证字符串和特殊验证字符串。基于文本相关识别技术得到第一嵌入码,第一嵌入码可由一般字符串数字串文本相关判别器提取,基于文本无关识别技术得到第二嵌入码,第二嵌入码可由预设的特殊字符串文本相关声纹判别器进行提取。第一嵌入码和第二嵌入码为该用户的声纹初始嵌入码。另外,在后续实施中一般字符串数字串文本相关判别器和文本无关判别器仅为1个,特殊字符串文本相关声纹判别器则可预设多个,每一个对应一个中文词组或多个字母或英文词组,且可以进行自主学习增加。

[0078] 假设一个验证语料共M位字符,则可拆分为N位字符的一般验证字符串和M-N位字符的特殊验证字符串。一般字符串一般由N位数字构成,M-N特殊字符一般由字母或英文词组或汉字构成,如此设计可有效避免算法对不同信道判别的性能退化问题。

[0079] 参看图4,为方便演示所展示的一组验证语料,也可称之为待验证语料,不以其具体内容进行限定。具体地,每个验证语料长度为8位,稍短于前述的注册字符串,其中包含 $N=6$ 位字符的一般字符串和包含 $M-N=2$ 位特殊字符的特殊字符串(字母或英文词组或汉字)构成。验证语料1包含了所有初始注册库中的内容;验证语料2则包含了未知语料,即“拼搏”属于特殊字符串中的中文词组,位置处于句末;验证语料3包含的“AX”特殊字符串中的英文字母,位置处于句末;验证语料4中均为初始注册语料,“烦恼”属于特殊字符串中的中文词组,位置处于句中;验证语料5,包含了未知语料,即“可爱”属于特殊字符串中的中文词组,位置处于句首。

[0080] 完成上述准备工作后,即使用者可以通过系统给出的动态验证码从而进行身份验

证。

[0081] 首先,参看图1和图2,在步骤S2中,使用者验证时根据系统提供的动态验证输入一动态验证语料。使用者可通过不同的音频录取设备输入动态验证语料。输入后可对动态验证语料进行视觉上加噪,提高安全性和防窃性,然后对动态验证语料进行环境语音质量检测(包括SNR计算)判断环境噪音水平,若计算得到的信噪比过低,且低于预设阈值,则会提示调整重新输入动态验证语料。超过三次后本系统直接结束,具体多少次数可进行自主调整,而高于预设阈值的动态验证语料则会转变为wav格式,但不限于该单一格式,并对其进行切分。得到一般动态验证字符串和特殊动态验证字符串,并分别进行识别,可分为已注册的一般动态验证字符串、已注册的特殊动态验证字符串、未注册的特殊动态验证字符串。判断为一般动态验证字符串则进入步骤S3进行后续处理,判断为特殊动态验证字符串则进入步骤S4进入后续处理,其中,一般动态验证字符串由若干数字组成,特殊动态验证字符串为中文词语或英文字母或英文词组组成。

[0082] 假设本次所需要输入的动态验证语料为“81567903茶杯”,则切分后的数字“81567903”和汉语词组“茶杯”分别送入语音识别引擎进行识别,即对数字“81567903”验证跳转至步骤S3,对“茶杯”验证跳转至步骤S4。

[0083] 参看图1和图2,在本实施例中,还在步骤S2中,对动态验证字符串进行数字识别,若识别通过则进入数字预处理模块进行后续处理,否则认定为验证失败并停止验证。还会在这之后通过语速检测器进行语速验证,若语速太快则提示重新录入,若多次提示后无效则可进行关闭。对于一般字符串,由于待验证时产生的语料文本顺序是随机的,故需要进行切分调序,以形成与注册声纹库相同的一般字符串顺序,再基于文本相关识别中的端到端声纹识别方法进行验证。因此,在数字预处理模块对动态验证字符串进行一一切分,得到8个仅包含一个数字的数字段。然后对该8个数字段根据一般注册字符串的排列顺序进行重新排列,形成“01356789”(为了方便理解,假设这里的一般注册字符串为0123456789)。接着,进入重建数字串模块并根据先验规则调整补齐缺失语料(即2和4,随机从注册语句中信噪比达到阈值的相应数字进行填充)以此重建数字串,填充排序后,形成“0123456789”。

[0084] 参看图1和图2,在步骤S3中,将重建的数字串送入一般字符串数字串文本相关判别器中进行判别,生成得到第一验证嵌入码,并将第一验证嵌入码与第一嵌入码进行求解以获取数字串相似性,以进行声纹匹配,若数字串相似性超过预设阈值,则判定成功进入后续步骤,否则认定为验证失败并停止。判定成功后将数字串同时送入文本无关语料库和文本相关语料库进行保存,丰富该用户的语音数据。

[0085] 假设系统内已经形成了T组文本相关的语料库 $SL = \{RC_0, RS_1, RS_3, \dots, RS_{T-1}\}$ ,这里 $RC_0$ 为一般字符串,其数量为 $N_0$ 、长度 $Lrc_0$ ,而 $RS_1, RS_3, \dots, RS_{T-1}$ 为已注册的特殊字符串,其数量和长度分别为 $N_1$ 与 $Lrc_1$ 、 $N_2$ 与 $Lrc_2$ 、 $\dots$ 、 $N_{T-1}$ 、长度 $Lrc_{T-1}$ 。对于语料库中出现的RS和目前未注册的特殊字符串(URS),则分为两种类型处理。

[0086] 对于一般字符串,将其利用预处理技术进行预处理,进行切分调序后的对齐语料添加至既有一般注册语料 $RC_0 = \{RC_{0_1}, RC_{0_2}, \dots, RC_{0_{N_0}}\}$ 中,变为新的 $RC_0 = \{RC_{0_1}, RC_{0_2}, \dots, RC_{0_{N_0}}, RC_{0_{N_0+1}}\}$ ,再重新进入神经网络编码器,用于生成新的嵌入码,然后再与其余嵌入码计算后形成新的嵌入码。对于已注册特殊字符串,参考处理方式类似,后续不再赘述。将第一验证嵌入码和其余第一嵌入码计算后形成新的第一嵌入码。

[0087] 参看图1和图2,步骤S4对特殊字符串进行判断。

[0088] 在步骤S2中,会对切分后的特殊动态验证字符串进行识别,若为同音则进入步骤S4,否则认定为验证失败并停止。接着,在步骤S4中,判断特殊动态验证字符串是否存在对应的特殊字符串文本相关声纹判别器即“茶杯”判别器,此处识别为系统在给出动态验证语料已经做好判断是否存在对应的特殊字符串文本相关声纹判别器,以及文本相关语料库是否存在待验证者特殊动态验证字符串的至少1条注册语料。若两者均存在则进入步骤S5,若相对应的特殊字符串文本相关声纹判别器不存在或两者均不存在则进入步骤S6,若相对应的特殊字符串文本相关声纹判别器存在但未激活以及文本相关语料库内至少1条注册语料存在,则进入步骤S7。

[0089] 参看图1和图2,在步骤S5中,将该特殊动态验证字符串“茶杯”送入相对应的特殊字符串文本相关声纹判别器进行判别,生成得到第二验证嵌入码,并将第二验证嵌入码与第二嵌入码进行求解以获取特殊动态验证字符串的相似性,以进行声纹匹配,若特殊动态验证字符串的相似性超过预设阈值,则判定成功进入后续步骤,否则认定为验证失败并停止。在步骤S52中,将特殊动态验证字符串送入相对应说话人的文本相关语料库和文本无关语料库进行保存,丰富该说话人语音数据。将第二验证嵌入码和其余第二嵌入码计算后形成新的第二嵌入码。

[0090] 参看图1、图2和图5,对于未注册的特殊动态验证字符串,基于文本无关方法开展声纹识别,以语句级或帧级嵌入码为物理量与注册声纹库中的嵌入码比较,设置阈值较低第二门限,实现二元判定;结合语音识别引擎判定设置阈值较高的第三门限;根据是否同时满足第二门限和第三门限的要求,实现二元未注册字符串的第四门限的判定。此外,系统具有自动更新和自学习功能,在同时满足第一门限和第四门限后,则认定为整条待测语料通过识别系统的判定。若通过判定,则进行分类处理并保存。

[0091] 而具体在步骤S6中,将特殊动态验证字符串“茶杯”送入文本无关判别器,以语句级或帧级嵌入码为物理量与注册声纹库中的嵌入码比较,以进行声纹匹配,若满足阈值则继续,否则认定为验证失败并停止。对特殊动态验证字符串通过语音识别引擎判定,若满足阈值则进入下一步,否则认定为验证失败并停止;若验证者注册语料存在则直接进入下一步。接着,将特殊动态验证字符串送入文本无关语料库保存,建立新的文本无关声纹识别模型,为后续在文本相关物料库中建立“茶杯”这一文本相关分类器做好准备。生成特殊字符串相关语料送入临时语料库中,当遍历说话人和语料均达到一定数量阈值时,将特殊字符串相关语料送入文本相关语料库进行保存,并且可以构建新的文本相关声纹判别器(“茶杯”分类器,并编号为原先最后一文本相关声纹判别器的加上1),未达到则继续收集此语料音频数据。另外,在执行步骤S6期间以至少一种语料为一组,一组语料可以为一种也可以为五种,以一种语料为一组作为例子,待“茶杯”成功建立起文本相关声纹识别分类器后,才能在动态验证语料出现另一组新的无关词组。即对于未注册特殊字符串,将其添加至临时SL中设定为 $RS_T$ ,当语料库中其长度 $N_T$ 超过某设定的阈值时,将 $RS_T$ 放入SL中,形成 $SL'$ ,训练形成新的神经网络编码器 $Net_T$ ,可用于第二嵌入码生成。由此,可以将动态验证语料进行拆分后贴上标签,与声纹数据库的注册语料集重新训练,形成新的注册声纹库、第一嵌入码、第二嵌入码和相对应新的嵌入码。

[0092] 当然上述存在一特殊情况,即步骤S7。若“茶杯”一词在预设时,词语判别器已经形

成[系统自己知道]了,但仍未在文本相关语料库中形成了注册语料的情况。现在该情况的原因是在刚注册之时,例如:总计录取5条,其中3条为数字,剩余2条为词语;也就是说,2条词语注册语料不会包含所有的100个判别器组所需的注册语料,例如覆盖了其中5个,剩余95个的注册词语需要在后续的验证过程中获取,假设下的“茶杯”正是如此。此时需要“茶杯”先进入文本无关语料库,然后对文本相关语料库后处理,以此激活带该“茶杯”的判别器。后续拿这5个做统计处理[最常见的例如平均,也可以有其他的统计量融合出一个新的注册嵌入码]。有N个,就N个做统计处理,这也是动态训练的从而在注册嵌入码上不断增强。

[0093] 参看图5,较优地,在步骤A3展示了本实施例的自身更新与自我学习。

[0094] 对于数字串的相关模型,初始时已经满足FAR和FRR条件,所以模型一般不更新,每验证一次增加一个数据,当数据每积累到一个阈值时,例如增加性能10%且测试FAR FRR同时优于之前指标,则更新模型,生成得到新的第一嵌入码。

[0095] 当100个文本相关判别器中,涉及到spk中缺乏该词语注册语料的,会进行注册语料更新,模型也同样更新。根据特殊验证字符串保存于文本相关资料库中的录入语料人数量及其语料数量同时满足一定阈值,则更新文本相关语料库以及特殊字符串文本相关声纹判别器。对文本相关语料库,以特定验证人为数据组织方式,将其所有语料以对抗训练方式训练相关的声纹合成网络;以既有语料文本作为输入,形成音频作为扩增的语料文件,并添加噪声信号,形成文本相关语料库内新的语料文件。举例说明如下,假设新增的“茶杯”一词,这个前期未形成的判别器,满足一定的条件(例如10000个说话人已有3组重复的词语后),然后再利用3组重复的词语进行数据扩增及鲁棒提升的方法,训练以形成模型。待模型形成后,则会对一些后续验证人中缺乏该词语注册语料的,会进行该验证人注册语料更新。与“茶杯”相同情况的几个文本相关声纹判别器,可以逐个增加,也可以以5个为一批完成后再积累下一个5个。

[0096] 文本无关判别器,当文本无关数据库中的数据积累到一个比例,例如增加10%且测试FAR FRR同时优于之前指标,则进行更新,得到新的文本无关判别器,以及新的嵌入码。

[0097] 上面结合附图对本发明的实施方式作了详细说明,但是本发明并不限于上述实施方式。即使对本发明作出各种变化,倘若这些变化属于本发明权利要求及其等同技术的范围之内,则仍落入在本发明的保护范围之内。



声纹库 初始注册1	5	1	8	7	0	9	2	4	3	6
声纹库 初始注册2	5	1	8	7	0	9	2	4	3	6
声纹库 初始注册3	D	B	G	H	F	D	B	G	H	F
声纹库 初始注册4	梦	想	烦	恼	晴	天	无	奈	存	储
声纹库 初始注册5	梦	想	烦	恼	晴	天	无	奈	存	储

图3

待验证语料1	7	0	9	5	1	6	梦	想
待验证语料2	8	7	3	0	9	2	拼	搏
待验证语料3	8	5	3	0	9	2	A	X
待验证语料4	7	6	2	烦	恼	4	3	8
待验证语料5	可	爱	2	3	2	4	3	8

图4

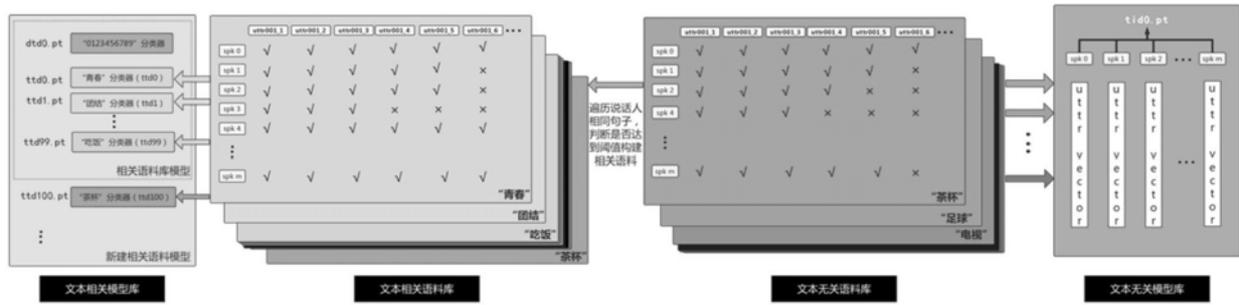


图5