(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114745720 A
(43)申请公布日 2022.07.12

(21)申请号 202210287790.0
(22)申请日 2022.03.23
(71)申请人 中国人民解放军战略支援部队信息工程大学
地址 450000 河南省郑州市高新区科学大道62号
(72)发明人 李邵梅 高超 黄瑞阳 朱宇航 王凯 李星 李英乐
(74)专利代理机构 郑州大通专利商标代理有限公司 41111
专利代理师 石丹丹
(51)Int.Cl.
H04W 12/12(2021.01)
H04W 12/128(2021.01)
G10L 25/27(2013.01)
G10L 25/51(2013.01)

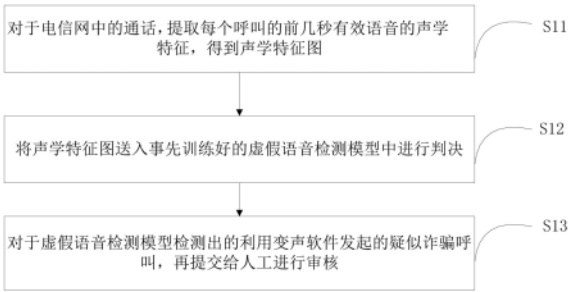
权利要求书1页 说明书4页 附图2页

(54)发明名称

变声型诈骗电话检测方法及装置

(57)摘要

本发明属于通信网内容安全检测技术领域，具体涉及一种变声型诈骗电话检测方法及装置，该方法包括：对于电信网中的通话，首先提取每个呼叫的前几秒有效语音的声学特征，得到声学特征图；然后将声学特征图送入虚假语音检测模型中进行判决，对于虚假语音检测模型检测出的利用变声软件发起的疑似诈骗呼叫，再提交给人工进行审核。本发明在检测过程中，不接触通话的具体内容，既不受诈骗分子更换话术内容的困扰，又能有效保护用户的通信隐私。



CN 114745720 A

1.一种变声型诈骗电话检测方法,其特征在于,包括:

对于电信网中的通话,首先提取每个呼叫的前几秒有效语音的声学特征,得到声学特征图;然后将声学特征图送入虚假语音检测模型中进行判决,对于虚假语音检测模型检测出的利用变声软件发起的疑似诈骗呼叫,再提交给人工进行审核。

2.根据权利要求1所述的变声型诈骗电话检测方法,其特征在于,所述声学特征采用美尔频率倒谱系数或者短时傅里叶变换对数幅度。

3.根据权利要求2所述的变声型诈骗电话检测方法,其特征在于,所述声学特征的提取过程如下:

电信网中语音的采样率是每秒8000个点,以256个采样点即32ms的语音为一帧,帧与帧之间有16ms的重叠,对于每个呼叫,采集主叫前4秒的语音内容进行处理,提取(4000-32)/16+1=249帧的声学特征参数;

美尔滤波器组中将语音在美尔频率上的分布划分成24个子带,即每帧语音经过美尔滤波器过滤后会得到24维的美尔频率倒谱系数。

4.根据权利要求3所述的变声型诈骗电话检测方法,其特征在于,对24维的美尔频率倒谱系数进行一阶和二阶差分,每帧语音得到72维的美尔频率倒谱系数;那么对于每个呼叫,得到249帧72维的声学特征参数。

5.根据权利要求4所述的变声型诈骗电话检测方法,其特征在于,用矩阵的形式把声学特征有序地组织起来,作为每个呼叫的声学特征图。

6.根据权利要求1所述的变声型诈骗电话检测方法,其特征在于,所述虚假语音检测模型为SVM分类模型、GMM分类模型或者深度神经网络模型。

7.根据权利要求6所述的变声型诈骗电话检测方法,其特征在于,所述深度神经网络模型采用CNN检测模型,所述CNN检测模型包括输入层、卷积层、池化层、全连接层和输出层。

8.根据权利要求7所述的变声型诈骗电话检测方法,其特征在于,所述输入层的尺寸为249*72;所述卷积层有3个5*5的卷积核,卷积时宽的步长是2,高的步长是1,填充的大小是1;所述池化层以2*2为单元,采用最大池化机制;所述输出层采用基于softmax的二分类输出。

9.根据权利要求7所述的变声型诈骗电话检测方法,其特征在于,所述CNN检测模型的训练过程如下:首先基于公开数据集中大量带标签的真实语音和虚假语音进行预训练,然后从电信网上采集少量的真实呼叫语音和利用变声软件发起的呼叫语音,人工进行标注后,再送入预训练模型中进行微调。

10.一种变声型诈骗电话检测装置,其特征在于,包括:

声学特征图提取模块,用于提取每个呼叫的前几秒有效语音的声学特征,得到声学特征图;

虚假语音检测模型判决模块,用于将声学特征图送入虚假语音检测模型中进行判决;

人工审核模块,用于对于虚假语音检测模型检测出的利用变声软件发起的疑似诈骗呼叫,再提交给人工进行审核。

变声型诈骗电话检测方法及装置

技术领域

[0001] 本发明属于通信网内容安全检测技术领域,具体涉及一种变声型诈骗电话检测方法及装置。

背景技术

[0002] 近年来,随着通信产业的迅猛发展,以电信网络诈骗为代表的网络诈骗成为危害人民群众财产安全的突出问题。为了实施诈骗,诈骗分子不断翻新手法,隐藏身份,更换话术剧本。随着以深度学习为代表的人工智能技术的发展,音频伪造生成技术不断成熟,基于合成和转换等技术生成的伪造语音的自然度和逼真度不断提升,已成为网络诈骗分子工具。在网络诈骗中,为了迷惑受害者,诈骗分子通常利用手机变声软件一人分饰多个角色,取得受害人的信任,进而达到骗取钱财的目的。

[0003] 从电信网海量呼叫中快速检测定位这些诈骗呼叫对于维护公民财产安全和国家安全稳定具有重要的意义。但是为了逃避基于名单的检测方法,诈骗分子通常会频繁更换号码;为了逃避基于内容的检测方法,诈骗分子通常会频繁地更换话术剧本。所以,挖掘号码和通话内容之外,这些诈骗呼叫更本质的特征对于诈骗电话的监管具有重要的意义。

发明内容

[0004] 针对利用变声软件发起的诈骗呼叫,本发明提出一种变声型诈骗电话检测方法及装置,不接触通话的具体内容,既不受诈骗分子更换话术内容的困扰,又能有效保护用户的通信隐私。

[0005] 为解决上述技术问题,本发明采用以下的技术方案:

[0006] 本发明提供了一种变声型诈骗电话检测方法,包括:

[0007] 对于电信网中的通话,首先提取每个呼叫的前几秒有效语音的声学特征,得到声学特征图;然后将声学特征图送入虚假语音检测模型中进行判决,对于虚假语音检测模型检测出的利用变声软件发起的疑似诈骗呼叫,再提交给人工进行审核。

[0008] 进一步地,所述声学特征采用美尔频率倒谱系数或者短时傅里叶变换对数幅度。

[0009] 进一步地,所述声学特征的提取过程如下:

[0010] 电信网中语音的采样率是每秒8000个点,以256个采样点即32ms的语音为一帧,帧与帧之间有16ms的重叠,对于每个呼叫,采集主叫前4秒的语音内容进行处理,提取(4000-32)/16+1=249帧的声学特征参数;

[0011] 美尔滤波器组中将语音在美尔频率上的分布划分成24个子带,即每帧语音经过美尔滤波器过滤后会得到24维的美尔频率倒谱系数。

[0012] 进一步地,对24维的美尔频率倒谱系数进行一阶和二阶差分,每帧语音得到72维的美尔频率倒谱系数;那么对于每个呼叫,得到249帧72维的声学特征参数。

[0013] 进一步地,用矩阵的形式把声学特征有序地组织起来,作为每个呼叫的声学特征图。

[0014] 进一步地，所述虚假语音检测模型为SVM分类模型、GMM分类模型或者深度神经网络模型。

[0015] 进一步地，所述深度神经网络模型采用CNN检测模型，所述CNN检测模型包括输入层、卷积层、池化层、全连接层和输出层。

[0016] 进一步地，所述输入层的尺寸为249*72；所述卷积层有3个5*5的卷积核，卷积时宽的步长是2，高的步长是1，填充的大小是1；所述池化层以2*2为单元，采用最大池化机制；所述输出层采用基于softmax的二分类输出。

[0017] 进一步地，所述CNN检测模型的训练过程如下：首先基于公开数据集中大量带标签的真实语音和虚假语音进行预训练，然后从电信网上采集少量的真实呼叫语音和利用变声软件发起的呼叫语音，人工进行标注后，再送入预训练模型中进行微调。

[0018] 本发明还提供了一种变声型诈骗电话检测装置，包括：

[0019] 声学特征图提取模块，用于提取每个呼叫的前几秒有效语音的声学特征，得到声学特征图；

[0020] 虚假语音检测模型判决模块，用于将声学特征图送入虚假语音检测模型中进行判决；

[0021] 人工审核模块，用于对于虚假语音检测模型检测出的利用变声软件发起的疑似诈骗呼叫，再提交给人工进行审核。

[0022] 与现有技术相比，本发明具有以下优点：

[0023] 本发明的变声型诈骗电话检测方法，对于电信网中的海量呼叫，逐个采用虚假语音检测技术对其通话语音进行判别，首先提取每个通话的声学特征，得到声学特征图，再采用虚假语音检测模型对其进行判决，被判为虚假语音的呼叫提交给人工进行进一步审核，在该检测过程中，不接触通话的具体内容，既不受诈骗分子更换话术内容的困扰，又能有效保护用户的通信隐私。

附图说明

[0024] 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，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，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。

[0025] 图1是本发明实施例的变声型诈骗电话检测方法的流程示意图；

[0026] 图2是本发明实施例的声学特征采用美尔频率倒谱系数的提取流程图；

[0027] 图3是本发明实施例的每个通话的美尔频率倒谱系数组成的声学特征图；

[0028] 图4是本发明实施例的CNN检测模块的结构图；

[0029] 图5是本发明实施例的变声型诈骗电话检测装置的结构框图，51表示声学特征图提取模块，52表示虚假语音检测模型判决模块，53表示人工审核模块。

具体实施方式

[0030] 为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是

本发明一部分实施例,而不是全部的实施例,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。

[0031] 如图1所示,本实施例的变声型诈骗电话检测方法,包含以下步骤:

[0032] 步骤S11,对于电信网中的通话,提取每个呼叫的前几秒有效语音的声学特征,得到声学特征图。

[0033] 步骤S12,将声学特征图送入事先训练好的虚假语音检测模型中进行判决。

[0034] 步骤S13,对于虚假语音检测模型检测出的利用变声软件发起的疑似诈骗呼叫,再提交给人工进行审核。

[0035] 为了兼顾检测的精度和效率,基于多次实验的结果,对于每个呼叫,我们只取前4秒的有效语音进行检测。

[0036] 作为优选的,所述声学特征可以采用Mel-Frequency Cepstral Coefficients(MFCCs,美尔频率倒谱系数),或者the Logarithmic Magnitude of Short-Time Fourier Transform(log-magnitude STFT,短时傅里叶变换对数幅度)等语音信号处理领域的声学参数。

[0037] 以声学特征采用MFCCs为例,声学特征的提取过程如下:

[0038] 如图2所示,图中的分帧过程是将一段连续的语音切分成多个小段来分别进行后续的处理,每个小段称之为一帧。电信网中语音的采样率是每秒8000个点,根据应用经验,以256个采样点即32ms的语音为一帧,按照语音信号处理的惯例,帧与帧之间有16ms的重叠,对于每个呼叫,采集主叫前4秒的语音内容进行处理,那么对于每个呼叫,可以提取(4000-32)/16+1=249帧的声学特征参数。

[0039] 另外,针对电话信道的语音,美尔滤波器组中将语音在美尔频率上的分布划分成24个子带,即每帧语音经过美尔滤波器过滤后会得到24维的美尔频率倒谱系数。为了提高声学特征刻画的精度,对初始的24维美尔频率倒谱系数再进行一阶和二阶差分,最终每帧语音得到72维的美尔频率倒谱系数(MFCCs);那么对于每个呼叫,得到249帧72维的声学特征参数,用矩阵的形式把这些声学特征有序地组织起来,可以作为每个呼叫的声学特征图(如图3所示)。

[0040] 所述虚假语音检测模型是一个二分类模型,可以用SVM(Support Vector Machine,支持向量机)、GMM(Gaussian Mixed Mode,高斯混合模型)等传统的分类模型,也可以用深度神经网络模型。在本实例中,所述虚假语音检测模型采用CNN(Convolutional Neural Networks,卷积神经网络)检测模型,把从每个呼叫语音提取出的声学特征转换成图的形式,然后利用CNN检测模型进行分类。

[0041] 如图4所示,所述CNN检测模型包括输入层、卷积层、池化层、全连接层和输出层;输入层的尺寸为249*72,卷积层有3个5*5的卷积核,卷积时宽的步长是2,高的步长是1,填充的大小是1,那么经过卷积后:

[0042] 特征图的宽度为:(249-5+2*1)/2+1=124;

[0043] 特征图的高度为:(72-5+2*1)/1+1=70;

[0044] 卷积层的输出维度为:124*70*3=26040;

[0045] 池化层以2*2为单元,采用最大池化机制,经过池化层后的输出维度为62*35*3=6510。全连接层的输入维度为6510,输出维度为1024,输出层采用基于softmax的二分类输

出。

[0046] 具体的,CNN检测模型的训练过程如下:

[0047] 首先采用ASVspoof2019竞赛中LA部分的训练数据,按照如图2和图3所示的过程提取声学特征图对CNN检测模型进行预训练,训练时,采用随机梯度下降的模型参数求解法,训练的批次大小batch_size=32,轮次epcho为300,得到可用于虚假语音检测的预训练模型;然后再采集部分通信网中正常呼叫的主叫语音和利用变声软件的虚假语音,人工进行标注后,同样提取声学特征图送入上述预训练模型中进行微调,微调的数据批次大小为32,轮次epcho为50,得到可用于虚假语音检测的CNN检测模型。

[0048] 与上述变声型诈骗电话检测方法相应地,如图5所示,本实施例还提供一种变声型诈骗电话检测装置,包括:

[0049] 声学特征图提取模块51,用于提取每个呼叫的前几秒有效语音的声学特征,得到声学特征图。

[0050] 虚假语音检测模型判决模块52,用于将声学特征图送入虚假语音检测模型中进行判决。

[0051] 人工审核模块53,用于对于虚假语音检测模型检测出的利用变声软件发起的疑似诈骗呼叫,再提交给人工进行审核。

[0052] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。

[0053] 最后需要说明的是:以上所述仅为本发明的较佳实施例,仅用于说明本发明的技术方案,并非用于限定本发明的保护范围。凡在本发明的精神和原则之内所做的任何修改、等同替换、改进等,均包含在本发明的保护范围内。

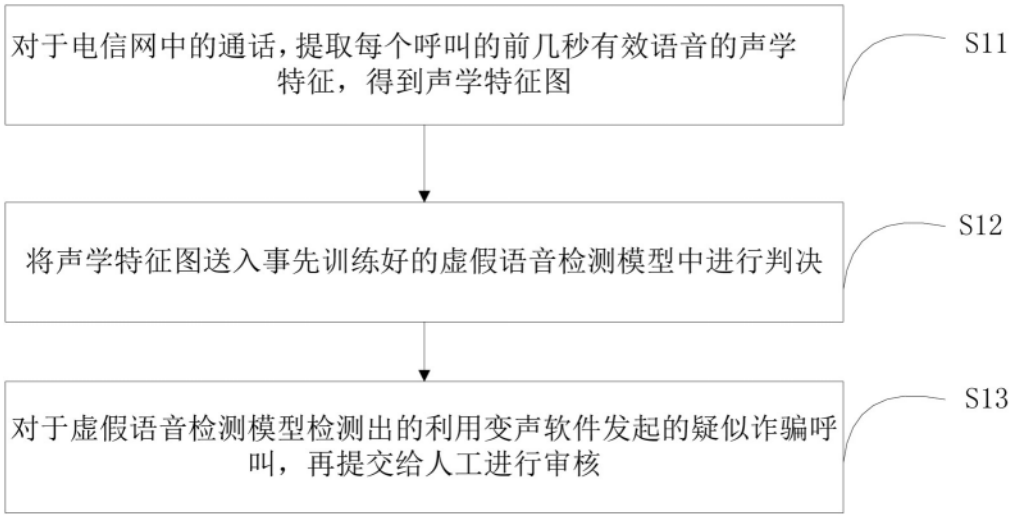

对于电信网中的通话，提取每个呼叫的前几秒有效语音的声学特征，得到声学特征图
S11
将声学特征图送入事先训练好的虚假语音检测模型中进行判决
S12
对于虚假语音检测模型检测出的利用变声软件发起的疑似诈骗呼叫，再提交给人工进行审核
S13

图1

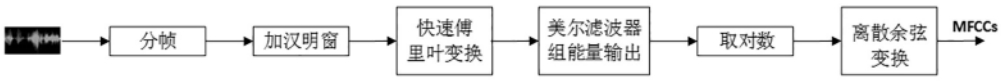

分帧
加汉明窗
快速傅里叶变换
美尔滤波器组能量输出
取对数
离散余弦变换
MFCCs

图2

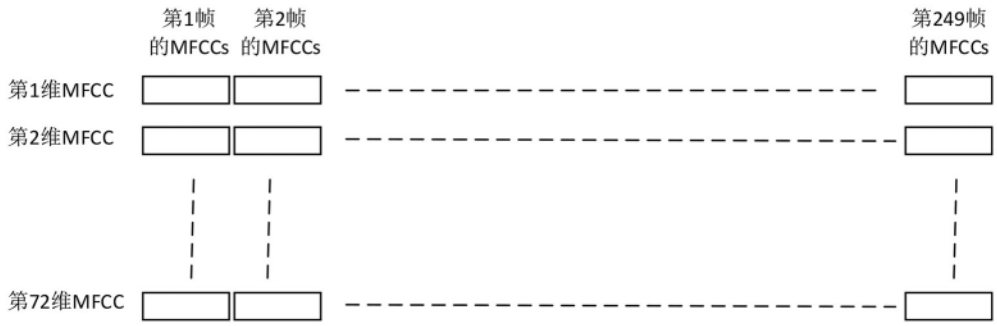

第1帧的MFCCs
第2帧的MFCCs
第249帧的MFCCs
第1维MFCC
第2维MFCC
第72维MFCC

图3

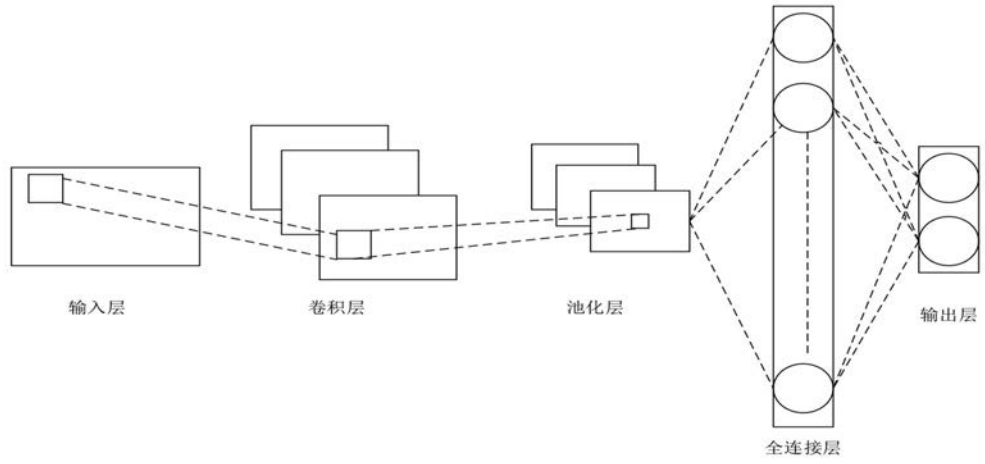


图4

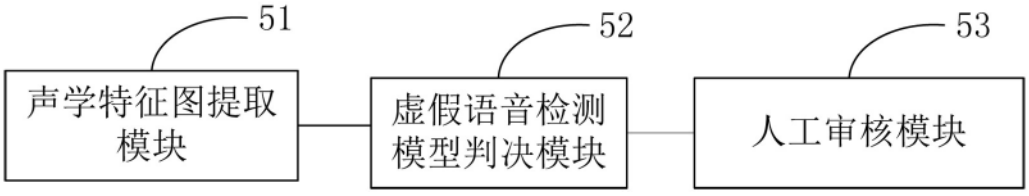


图5