



(12) 发明专利

(10) 授权公告号 CN 110751944 B

(45) 授权公告日 2024.09.24

(21) 申请号 201910884620.9

G10L 15/16 (2006.01)

(22) 申请日 2019.09.19

G10L 15/26 (2006.01)

G10L 25/18 (2013.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 110751944 A

(43) 申请公布日 2020.02.04

(73) 专利权人 平安科技(深圳)有限公司

地址 518033 广东省深圳市福田区福田街
道福安社区益田路5033号平安金融中
心23楼

(56) 对比文件

CN 108847223 A, 2018.11.20

CN 109767759 A, 2019.05.17

CN 108694951 A, 2018.10.23

ZHANG XIANGYU.elt.Deep Residual Learning for Image Recognition.2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).2016,P 770 - 778.

(72) 发明人 王健宗 贾雪丽

审查员 谭宇玲

(74) 专利代理机构 北京市京大律师事务所

11321

专利代理师 谢文强

(51) Int. Cl.

G10L 15/06 (2013.01)

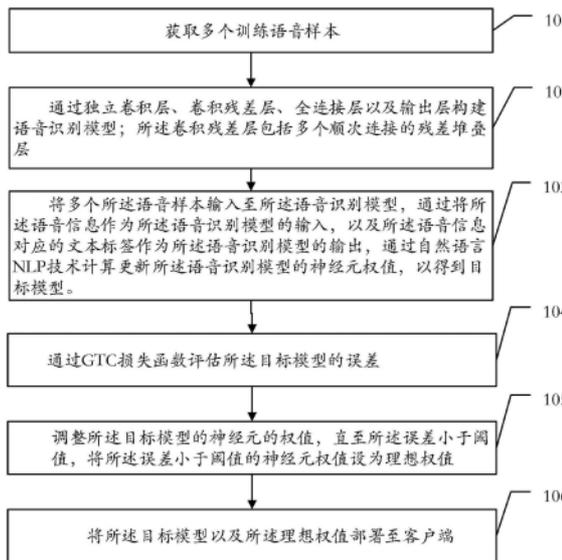
权利要求书3页 说明书12页 附图2页

(54) 发明名称

构建语音识别模型的方法、装置、设备和存储介质

(57) 摘要

本申请涉及人工智能领域,提供了一种构建语音识别模型的方法、装置、设备和存储介质,方法包括:获取多个训练语音样本;通过独立卷积层、卷积残差层、全连接层以及输出层构建语音识别模型;将所述训练语音信息输入至所述语音识别模型,通过自然语言处理NLP技术、所述语音信息以及所述语音信息对应的文本标签更新所述语音识别模型的神经元权值,以得到目标模型;通过L(S) = -lnΠ_{(h(x),z)∈S} p(z|h(x)) = -Σ_{(h(x),z)∈S} ln p(z|h(x)) 评估所述目标模型的误差;调整所述目标模型的神经元的权值,直至所述误差小于阈值,将所述误差小于阈值的神经元权值设为理想权值;将所述目标模型以及所述理想权值部署至客户端。降低了语音信息中的音调对预测文本的影响以及语音识别模型识别过程中的运算量。



1. 一种构建语音识别模型的方法,其特征在于,所述方法包括:

获取多个训练语音样本,所述训练语音样本包括语音信息以及与语音信息对应的文本标签;

通过独立卷积层、卷积残差层、全连接层以及输出层构建语音识别模型,所述卷积残差层包括多个顺次连接的残差堆叠层,所述残差堆叠层包含多个顺次连接的残差模块,所述残差模块包含多个顺次连接的隐藏层以及旁路于多个顺次连接的权值层的旁路通道;所述独立卷积层用于从所述语音信息中提取声学特征,并消除所述声学特征中的非极大值;所述声学特征包括音节的发音、用户连读习惯以及语音频谱;所述卷积残差层用于将所述声学特征映射到隐层特征空间;所述全连接层用于整合映射到所述隐层特征空间的声学特征,以获取所述声学特征的含义,根据所述含义输出各种文本类型所对应的概率;所述输出层用于根据所述各种文本类型对应的概率输出所述语音信息所对应的文本;

将多个所述语音样本依次输入至所述语音识别模型,将所述语音信息及将所述语音信息对应的文本标签分别作为所述语音识别模型的输入以及输出,通过所述输入以及所述输出不断训练所述语音识别模型的神经元权值,直至所述语音样本均已输入至所述语音识别模型,结束对所述语音识别模型的训练,所述训练结束后,将带有训练好神经元权值的所述语音识别模型作为目标模型;

通过 $L(S) = -\ln \prod_{(h(x), z) \in S} p(z | h(x)) = -\sum_{(h(x), z) \in S} \ln p(z | h(x))$ 评估所述目标模型的误差,其中, $L(S)$ 为所述误差, x 为所述语音信息, z 为所述文本标签, $p(z | h(x))$ 为预测文本与所述文本标签的相似度, S 为所述多个训练语音样本,所述预测文本是指所述语音信息输入至所述目标模型后,由所述目标模型根据神经元权值计算输出的文本信息;

调整所述目标模型的神经元的权值,直至所述误差小于阈值,将所述误差小于阈值的神经元权值设为理想权值;

将所述目标模型以及所述理想权值部署至客户端;

所述全连接层包括分类函数,所述分类函数是指 $\delta(z)_j = \frac{\exp(z_j)}{\sum_{k=1}^K \exp(z_k)}$,所述 j 为自然数,

所述分类函数将卷积残差层输出的 K 维的语音频域信号向量 z 压缩到另一个 K 维实向量 $\delta(z)_j$,使得每一个元素的范围都在 $(0, 1)$ 之间,并且所有元素的和为1;其中,所述元素为 K 维实向量 $\delta(z)_j, j=1, \dots, K$;

所述残差模块的输入为 x ,所述残差模块的输出为 y ,所述残差模块的数学表达式为:

$y = F(x, w_i) + w_s x$,所述 $F(x, w_i)$ 为所述独立卷积层的输出,所述 w_s 为所述残差模块的权值, w_i 为所述残差模块的所有权值。

2. 根据权利要求1所述的方法,其特征在于,所述将多个所述语音样本输入至所述语音识别模型之前,所述方法还包括:

根据预设的分帧参数分帧处理训练语音信息,得到与所述训练语音信息对应的语句,所述预设分帧参数包括帧时长、帧数和前后帧重复时长;

根据预设的二维参数和滤波器组的特征提取转化为所述语句,得到二维语音信息。

3. 根据权利要求2所述的方法,其特征在于,所述根据预设的分帧参数分帧处理训练语音信息,包括:

对所述二维语音信息进行离散傅里叶变换,以得到所述二维语音信息对应的线性频谱 $X(k)$;

通过预设的带通滤波器对所述线性频谱滤波,以得到目标线性频谱,当所述带通滤波器的中心频率为 $f(m)$ 时,则所述带通滤波器的传递函数为:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases}, \text{所述 } f(m) \text{ 的表达式为:}$$

$$f(m) = \left(\frac{N}{f_s} F_{mel}^{-1}(F_{mel}(f_l)) + m \frac{F_{mel}(f_h) - F_{mel}(f_l)}{M+1} \right),$$

所述带通滤波器包括多个具有三角形滤波特性的带通滤波器,所述 f_l 为所述带通滤波器频率范围的最低频率,所述 f_h 为所述带通滤波器频率范围的最高频率,所述 N 为离散傅里叶变换时的长度,所述 f_s 为所述带通滤波器的采样频率,所述 F_{mel} 函数为 $F_{mel} = 1125 \ln(1+f/70)$,所述 F_{mel} 的逆函数为: $F_{mel}^{-1}(b) = 700(e^{b/1125} - 1)$, b 为整数;

根据 $s(m) = \ln(\sum_{k=0}^{N-1} |X(k)|^2 |H_m(k)|)$, $0 \leq m \leq M$ 计算所述目标线性频谱对应的对数能量,得到语谱图,所述 $X(k)$ 为所述线性频谱。

4. 根据权利要求1所述的方法,其特征在于,所述 $F(x, w_i)$ 的采用ReLU函数作为所述独立卷积层的激活函数,所述ReLU函数的数学表达式为 $\text{ReLU}(x) = \max(0, x)$ 。

5. 根据权利要求1所述的方法,其特征在于,所述调整所述目标模型的神经元的权值,包括:

通过随机梯度下降法调整所述神经元的权值。

6. 一种构建语音识别模型的装置,其特征在于,所述装置包括:

输入输出模块,用于获取多个训练语音样本,所述训练语音样本包括语音信息以及与语音信息对应的文本标签;

处理模块,用于通过独立卷积层、卷积残差层、全连接层以及输出层构建语音识别模型,所述卷积残差层包括多个顺次连接的残差堆叠层,所述残差堆叠层包含多个顺次连接的残差模块,所述残差模块包含多个顺次连接的隐藏层以及旁路于多个顺次连接的权值层的旁路通道;所述独立卷积层用于从所述语音信息中提取声学特征,并消除所述声学特征中的非极大值;所述声学特征包括音节的发音、用户连读习惯以及语音频谱;所述卷积残差层用于将所述声学特征映射到隐层特征空间;所述全连接层用于整合映射到所述隐层特征空间的声学特征,以获取所述声学特征的含义,根据所述含义输出各种文本类型所对应的概率;所述输出层用于根据所述各种文本类型对应的概率输出所述语音信息所对应的文本;通过输入输出模块将多个所述语音样本依次输入至所述语音识别模型,将所述语音信息及将所述语音信息对应的文本标签分别作为所述语音识别模型的输入以及输出,通过所述输入以及所述输出不断训练所述语音识别模型的神经元权值,直至所述语音样本均已输入至所述语音识别模型,结束对所述语音识别模型的训练;所述训练结束后,将带有训练好

神经元权值的所述语音识别模型作为目标模型;通过 $L(S) = -\ln \prod_{(h(x), z) \in S} p(z|h(x)) = -\sum_{(h(x), z) \in S} \ln p(z|h(x))$ 评估所述目标模型的误差;其中, $L(S)$ 为所述误差, x 为所述语音信息, z 为所述文本标签, $p(z|h(x))$ 为预测文本与所述文本标签的相似度, S 为所述多个训练语音样本;所述预测文本是指所述语音信息输入至所述目标模型后,由所述目标模型根据神经元权值计算输出的文本信息;

调整所述目标模型的神经元的权值,直至所述误差小于阈值,将所述误差小于阈值的神经元权值设为理想权值;将所述目标模型以及所述理想权值部署至客户端;

所述全连接层包括分类函数,所述分类函数是指 $\delta(z)_j = \frac{\exp(z_j)}{\sum_{k=1}^K \exp(z_k)}$,所述 j 为自然数,

所述分类函数将卷积残差层输出的 K 维的语音频域信号向量 z 压缩到另一个 K 维实向量 $\delta(z)_j$,使得每一个元素的范围都在 $(0, 1)$ 之间,并且所有元素的和为1;其中,所述元素为 K 维实向量 $\delta(z)_j$, $j=1, \dots, K$;

所述残差模块的输入为 x ,所述输出残差模块的输出为 y ,所述残差模块的数学表达式为:

$y = F(x, w_i) + w_s x$,所述 $F(x, w_i)$ 为所述独立卷积层的输出,所述 w_s 为所述残差模块的权值, w_i 为所述残差模块的所有权值。

7. 根据权利要求6所述的装置,其特征在于,所述处理模块还用于:

根据预设的分帧参数分帧处理训练语音信息,得到与所述训练语音信息对应的语句,所述预设分帧参数包括帧时长、帧数和前后帧重复时长;

根据预设的二维参数和滤波器组的特征提取转化为所述语句,得到二维语音信息。

8. 一种构建语音识别模型的设备,其特征在于,所述构建语音识别模型的设备包括:至少一个处理器、存储器和输入输出单元;

其中,所述存储器用于存储程序代码,所述处理器用于调用所述存储器中存储的程序代码来执行如权利要求1-5中任一项所述的方法。

9. 一种计算机存储介质,其特征在于,其包括指令,当其在计算机上运行时,使得计算机执行如权利要求1-5中任一项所述的方法。

构建语音识别模型的方法、装置、设备和存储介质

技术领域

[0001] 本申请涉及智能决策领域,尤其涉及一种构建语音识别模型的方法、装置、设备和存储介质。

背景技术

[0002] 语音识别用于将语音转换为文本。随着深度学习技术的不断发展,语音识别的应用范围也越来越广。

[0003] 目前,深度神经网络(deep neural networks,DNN)已经成为自动语音识别领域研究的热点。卷积神经网络(convolutional neural networks,CNN)、循环神经网络(recurrent neural networks,RNN)在语音识别模型创建上都取得了比较好的效果,深度学习已经成为语音识别的主流方案。

[0004] 在深度神经网络中,网络的深度往往与识别的正确率密切相关,因为传统的深度神经网络能够提取到低层、中层以及高层(low/mid/high-level)的多层次特征,网络的层数越多,意味着提取到的特征越丰富。但是,随着网络层级的不断加深,深度神经网络的“退化现象”也开始出现,导致语音识别的准确率很快达到饱和,出现网络层级越深,错误率反而越高的现象。此外,现有的语音识别模型在训练之前需要对语音训练样本进行对齐操作,对每一帧的语音数据与对应的标签进行对齐,以保证训练中所使用的损失函数能够准确估计语音识别模型的训练误差。然而,语音训练样本的对齐过程繁琐、复杂,需要耗费很大的时间成本。

发明内容

[0005] 本发明实例中通过获取无标注数据的特征,将获得的特征其引入到监督学习中,使得可使用的样本数据得到扩充和提高未标注图像的利用效率,模型预测的准确率上升。

[0006] 第一方面,本申请提供一种构建语音识别模型方法,包括:

[0007] 获取多个训练语音样本,所述训练语音样本包括语音信息以及与语音信息对应的文本标签;

[0008] 通过独立卷积层、卷积残差层、全连接层以及输出层构建语音识别模型,所述卷积残差层包括多个顺次连接的残差堆叠层,所述残差堆叠层包含多个顺次连接的残差模块,所述残差模块包含多个顺次连接的隐藏层以及旁路于多个顺次连接的权值层的旁路通道;

[0009] 将多个所述语音样本依次输入至所述语音识别模型,将所述语音信息及将所述语音信息对应的文本标签分别作为所述语音识别模型的输入以及输出,通过所述输入以及所述输出不断训练所述语音识别模型的神经元权值,直至所述语音样本均已输入至所述语音识别模型,结束对所述语音识别模型的训练,所述训练结束后,将带有训练好神经元权值的所述语音识别模型作为目标模型;

[0010] 通过 $L(S) = -\ln \prod_{(h(x), z) \in S} p(z|h(x)) = -\sum_{(h(x), z) \in S} \ln p(z|h(x))$ 评估所述目标模型的误差,其中, $L(S)$ 为所述误差, x 为所述语音信息, z 为所述文本标签, $p(z|h(x))$ 为所

述预测文本与所述文本标签的相似度, S为所述多个训练语音样本, 所述预测文本是指所述语音信息输入至所述目标模型后, 由所述目标模型根据神经元权值计算输出的文本信息;

[0011] 调整所述目标模型的神经元的权值, 直至所述误差小于阈值, 将所述误差小于阈值的神经元权值设为理想权值;

[0012] 将所述目标模型以及所述理想权值部署至客户端。

[0013] 在一些可能的设计中, 所述将多个所述语音样本输入至所述语音识别模型之前, 所述方法还包括:

[0014] 根据预设的分帧参数分帧处理所述训练语音信息, 得到所述训练语音信息对应的语句, 所述预设分帧参数包括帧时长、帧数和前后帧重复时长;

[0015] 根据预设的二维参数和滤波器组特征提取算法转化所述语句, 得到二维语音信息。

[0016] 在一些可能的设计中, 所述根据预设的分帧参数分帧处理所述训练语音信息, 包括:

[0017] 对所述二维语音信息进行离散傅里叶变换, 以得到所述二维语音信息对应的线性频谱 $X(k)$;

[0018] 通过预设的带通滤波器对所述线性频谱滤波, 以得到目标线性频谱, 当所述带通滤波器的中心频率为 $f(m)$ 时, 则所述带通滤波器的传递函数为:

$$[0019] \quad H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad \text{所述 } f(m) \text{ 的表达式为:}$$

$$[0020] \quad f(m) = \left(\frac{N}{f_s} F_{mel}^{-1}(F_{mel}(f_l)) + m \frac{F_{mel}(f_h) - F_{mel}(f_l)}{M+1} \right),$$

[0021] 所述带通滤波器包括多个具有三角形滤波特性的带通滤波器, 所述 f_l 为所述带通滤波器频率范围的最低频率, 所述 f_h 为带所述通滤波器频率范围的最高频率, 所述 N 为DFT时的长度, 所述 f_s 为所述带通滤波器的采样频率, 所述 F_{mel} 函数为 $F_{mel} = 1125 \ln(1+f/700)$, 所述 F_{mel} 的逆函数为: $F_{mel}^{-1}(b) = 700(e^{b/1125} - 1)$, b 为整数;

[0022] 根据 $s(m) = \ln(\sum_{k=0}^{N-1} |X(k)|^2 H_m(k))$, $0 \leq m \leq M$ 计算所述目标线性频谱对应的对数能量, 得到语谱图, 所述 $X(k)$ 为所述线性频谱。

[0023] 在一些可能的设计中, 所述全连接层包括分类函数, 所述分类函数是指

$$\delta(z)_j = \frac{\exp(z_j)}{\sum_{k=1}^K \exp(z_k)}, \quad \text{所述 } j \text{ 为自然数, 所述分类函数将卷积残差层输出的 } K \text{ 维的语音频域信号向量 } z \text{ 压缩到另一个 } K \text{ 维实向量 } \delta(z)_j, \text{ 使得每一个元素的范围都在 } (0, 1) \text{ 之间, 并且所有元素的和为 } 1。$$

[0024] 在一些可能的设计中, 所述残差模块的输入为 x , 所述输出残差模块的输出为 y , 则

所述残差模块的数学表达式为：

[0025] $y = F(x, w_i) + w_s x$, 所述 $F(x, w_i)$ 为所述独立卷积层的输出, 所述 w_s 为所述残差模块的权值。

[0026] 在一些可能的设计中, 所述 $F(x, w_i)$ 的采用ReLU函数作为所述独立卷积层的激活函数, 所述ReLU函数的数学表达式为 $\text{ReLU}(x) = \max(0, x)$,

[0027] 在一些可能的设计中, 所述调整所述目标模型的神经元的权值, 包括:

[0028] 通过随机梯度下降法调整所述神经元的权值。

[0029] 第二方面, 本申请提供一种构建语音识别模型的装置, 具有实现对应于上述第一方面提供的构建语音识别模型的平台的功能。所述功能可以通过硬件实现, 也可以通过硬件执行相应的软件实现。硬件或软件包括一个或多个与上述功能相对应的模块, 所述模块可以是软件和/或硬件。

[0030] 所述构建语音识别模型的装置包括:

[0031] 获取模块, 用于获取多个训练语音样本, 所述训练语音样本包括语音信息以及与语音信息对应的文本标签;

[0032] 处理模块, 用于通过独立卷积层、卷积残差层、全连接层以及输出层构建语音识别模型, 所述卷积残差层包括多个顺次连接的残差堆叠层, 所述残差堆叠层包含多个顺次连接的残差模块, 所述残差模块包含多个顺次连接的隐藏层以及旁路于多个顺次连接的权值层的旁路通道, 通过输入输出模块将多个所述语音样本依次输入至所述语音识别模型, 将所述语音信息及将所述语音信息对应的文本标签分别作为所述语音识别模型的输入以及输出, 通过所述输入以及所述输出不断训练所述语音识别模型的神经元权值, 直至所述语音样本均已输入至所述语音识别模型, 结束对所述语音识别模型的训练, 所述训练结束后, 将带有训练好神经元权值的所述语音识别模型作为目标模型, 通过 $L(S) = -\ln \prod_{(h(x), z) \in S} p(z|h(x)) = -\sum_{(h(x), z) \in S} \ln p(z|h(x))$ 评估所述目标模型的误差, 其中, $L(S)$ 为所述误差, x 为所述语音信息, z 为所述文本标签, $p(z|h(x))$ 为所述预测文本与所述文本标签的相似度, S 为所述多个训练语音样本, 所述预测文本是指所述语音信息输入至所述目标模型后, 由所述目标模型根据神经元权值计算输出的文本信息;

[0033] 调整所述目标模型的神经元的权值, 直至所述误差小于阈值, 将所述误差小于阈值的神经元权值设为理想权值, 将所述目标模型以及所述理想权值部署至客户端。

[0034] 在一些可能的设计中, 所述处理模块还用于:

[0035] 根据预设的分帧参数分帧处理所述训练语音信息, 得到所述训练语音信息对应的语句, 所述预设分帧参数包括帧时长、帧数和前后帧重复时长;

[0036] 根据预设的二维参数和滤波器组特征提取算法转化所述语句, 得到二维语音信息。

[0037] 在一些可能的设计中, 所述处理模块还用于:

[0038] 对所述二维语音信息进行离散傅里叶变换, 以得到所述二维语音信息对应的线性频谱 $X(k)$;

[0039] 通过预设的带通滤波器对所述线性频谱滤波, 以得到目标线性频谱, 当所述带通滤波器的中心频率为 $f(m)$ 时, 则所述带通滤波器的传递函数为:

$$[0040] \quad H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases},$$

所述f(m)的表达式为:

$$[0041] \quad f(m) = \left(\frac{N}{f_s} \right) F_{mel}^{-1} \left(F_{mel}(f_l) + m \frac{F_{mel}(f_h) - F_{mel}(f_l)}{M+1} \right),$$

[0042] 所述带通滤波器包括多个具有三角形滤波特性的带通滤波器,所述 f_l 为所述带通滤波器频率范围的最低频率,所述 f_h 为所述带通滤波器频率范围的最高频率,所述N为DFT时的长度,所述 f_s 为所述带通滤波器的采样频率,所述 F_{mel} 函数为 $F_{mel} = 1125 \ln(1+f/700)$,所述 F_{mel} 的逆函数为: $F_{mel}^{-1}(b) = 700(e^{b/1125} - 1)$,b为整数;

[0043] 根据 $s(m) = \ln(\sum_{k=0}^{N-1} |X(k)|^2 H_m(k))$, $0 \leq m \leq M$ 计算所述目标线性频谱对应的对数能量,得到语谱图,所述 $X(k)$ 为所述线性频谱。

[0044] 在一些可能的设计中,所述全连接层包括分类函数,所述分类函数是指

$$\delta(z)_j = \frac{\exp(z_j)}{\sum_{k=1}^K \exp(z_k)},$$

所述j为自然数,所述分类函数将卷积残差层输出的K维的语音频域信号向量z压缩到另一个K维实向量 $\delta(z)_j$,使得每一个元素的范围都在(0,1)之间,并且所有元素的和为1。

[0045] 在一些可能的设计中,所述处理模块还用于:所述残差模块的输入为x,所述输出残差模块的输出为y,则所述残差模块的数学表达式为:

$$[0046] \quad y = F(x, w_i) + w_s x,$$

所述 $F(x, w_i)$ 为所述独立卷积层的输出,所述 w_s 为所述残差模块的权值。

[0047] 在一些可能的设计中,所述 $F(x, w_i)$ 的采用ReLU函数作为所述独立卷积层的激活函数,所述ReLU函数的数学表达式为 $\text{ReLU}(x) = \max(0, x)$ 。

[0048] 在一些可能的设计中,所述调整所述目标模型的神经元的权值,包括:

[0049] 通过随机梯度下降法调整所述神经元的权值。

[0050] 本申请又一方面提供了一种构建语音识别模型的设备,其包括至少一个连接的处理器、存储器、输入输出单元,其中,所述存储器用于存储程序代码,所述处理器用于调用所述存储器中的程序代码来执行上述各方面所述的方法。

[0051] 本申请又一方面提供了一种计算机存储介质,其包括指令,当其在计算机上运行时,使得计算机执行上述各方面所述的方法。

[0052] 本申请通过旁路通道将输入信息x直接绕道传至隐含层的输出,旁路通道没有权值,保护了输入信息x的完整性,使得神经网络训练的更深,整个神经网络只需要训练输入、输出差别的部分,即传递输入信息x后,每个残差模块只学习残差 $F(x)$,简化训练的目标和难度,且神经网络稳定易于训练,随着神经网络深度的增加,语音识别模型的性能也逐渐变好,并且以CTC损失函数评估语音识别模型的预测文本,无需考虑文本标签中的发音音素与

训练语音信息的序列之间精准的映射关系,只需要输入序列和输出序列即可训练语音识别模型,节省了训练语音样本集的制作成本。此外,采用三角带通滤波器对所述训练语音信息的频谱进行平滑,消除所述训练语音信息中的谐波,突显原始声音的共振峰,避免语音信息中的音调对语音识别模型预测文本的影响,并降低语音识别模型识别过程中的对语音信息的运算量。

附图说明

- [0053] 图1为本申请实施例中构建语音识别模型的方法的流程示意图;
- [0054] 图2为本申请实施例中构建语音识别模型的装置的结构示意图;
- [0055] 图3为本申请实施例中构建语音识别模型的设备的结构示意图。
- [0056] 本申请目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

[0057] 应当理解,此处所描述的具体实施例仅用以解释本申请,并不用于限定本申请。本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的实施例能够以除了在这里图示或描述的内容以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或模块的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或模块,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或模块,本申请中所出现的模块的划分,仅仅是一种逻辑上的划分,实际应用中实现时可以有另外的划分方式,例如多个模块可以结合成或集成在另一个系统中,或一些特征可以忽略,或不执行。

[0058] 为解决上述技术问题,本申请主要提供以下技术方案:

[0059] 本申请通过旁路通道将输入信息 x 直接绕道传至隐含层的输出,旁路通道没有权值,保护了输入信息 x 的完整性,使得神经网络训练的更深,整个神经网络只需要训练输入、输出差别的部分,即传递输入信息 x 后,每个残差模块只学习残差 $F(x)$,简化训练的目标和难度,且神经网络稳定易于训练,随着神经网络深度的增加,语音识别模型的性能也逐渐变好,并且以CTC损失函数评估语音识别模型的预测文本,无需考虑文本标签中的发音音素与训练语音信息的序列之间精准的映射关系,只需要输入序列和输出序列即可训练语音识别模型,节省了训练语音样本集的制作成本。此外,采用三角带通滤波器对训练语音信息的频谱进行平滑,消除训练语音信息中的谐波,突显原始声音的共振峰,避免语音信息中的音调对语音识别模型预测文本的影响,并降低语音识别模型识别过程中的对语音信息的运算量。

[0060] 请参照图1,以下对本申请提供一种构建语音识别模型的方法进行举例说明,方法包括:

- [0061] 101、获取多个训练语音样本。
- [0062] 训练语音样本包括语音信息以及与语音信息对应的文本标签。
- [0063] 文本标签用于标注训练语音信息的发音音素。

[0064] 所述语音信息根据预先录制语音,将录音的内容写成文本;按照词语的先后顺序,对文本中的词语进行编号,对每个词语根据其发音音素进行标注,得到文本标签。文本标签中每个发音音素对应于录音中的一帧或多帧数据。

[0065] 102、通过独立卷积层、卷积残差层、全连接层以及输出层构建语音识别模型。

[0066] 卷积残差层包括多个顺次连接的残差堆叠层。残差堆叠层包含多个顺次连接的残差模块。残差模块包含多个顺次连接的隐藏层以及旁路于多个顺次连接的权值层的旁路通道。

[0067] 独立卷积层用于从语音信息中提取声学特征,并消除声学特征中的非极大值,降低声学特征的复杂度。声学特征包括特定音节的发音、用户连读习惯以及语音频谱等。

[0068] 卷积残差层用于将声学特征映射到隐层特征空间。

[0069] 全连接层用于整合映射到隐层特征空间的声学特征,以获取声学特征的含义,根据含义输出各种文本类型所对应的概率。

[0070] 输出层用于根据各种文本类型所对应的概率输出语音信息所对应的文本。

[0071] 本实施例中的语音识别模型为若干个顺次连接的隐藏层增加了旁路通道,以解决传统的神经网络随着网络层数的增加,训练准确率越来越低的问题。语音识别模型的卷积残差层有很多旁路通道,旁路通道作为隐藏层的支线,实现隐藏层之间的跨层连接,即将隐藏层的输入直接连到下一级层,使得下一级层可以直接学习残差。

[0072] 具体地,如图2所示,在一个残差模块中,跨层连接一般只跨越2至3个隐藏层,但不排斥跨越更多的隐藏层。仅跨越1个隐藏层的情况意义不大,实验效果不理想。

[0073] 假定残差模块的输入为 x ,期望输出是 $H(x)$,即 $H(x)$ 是期望的复杂潜在映射,但通常 $H(x)$ 的学习难度很大;如果直接把输入 x 传到输出作为初始结果,那么此时残差模块需要学习的目标就是 $F(x) = H(x) - x$ 。于是,相对于传统的神经网络,本实施例中的语音识别模型相当于将学习目标改变了,不再是学习一个完整的输出,而是学习最优解 $H(X)$ 和全等映射 x 的差值,即残差: $F(x) = H(x) - x$ 。

[0074] 从整体功能上看,如果用 $\{w_i\}$ 表示残差模块的所有权值,那么残差模块实际上计算的输出结果为:

$$[0075] \quad y = F(x, \{w_i\}) + x$$

[0076] 以跨越2个隐藏层为例,在忽略偏置的情况下, $F(x, \{w_i\}) = w_2 \delta(w_1 x) = w_2 \text{ReLU}(w_1 x)$,其中,ReLU函数为残差模块的激活函数。

[0077] 可以理解的是, $F(x, \{w_i\})$ 与 x 需要具有相同的维数。如果它们的维数不相同,则可以引入一个额外的权值矩阵 w_s 对 x 进行线性投影,使得 $F(x, \{w_i\})$ 与 x 的维数相同,相应地,残差模块的计算结果为: $y = F(x, \{w_i\}) + w_s x$

[0078] 将多个语音样本依次输入至语音识别模型,将语音信息及将语音信息对应的文本标签分别作为语音识别模型的输入以及输出,通过输入以及输出不断训练语音识别模型的神经元权值,直至语音样本均已输入至语音识别模型,结束对语音识别模型的训练。训练结束后,将带有训练好神经元权值的语音识别模型作为目标模型。

[0079] 在训练过程中,随机初始化语音识别模型内部的神经元的权值,再以训练语音信息作为语音识别模型的输入,以训练语音信息对于的文本标签作为语音识别模型的输出参考。训练语音信息在语音识别模型中运行前向传播,语音识别模型利用各层初始化后的神

经元对训练语音信息进行随机分类,最终得到与训练语音信息对应的预测文本。然后根据语音识别模型输出的预测文本和文本标签之间的差距来更新神经元的权值,再继续下一轮迭代,直至神经元的权值逼近要求值。

[0080] 103、通过 $L(S) = -\ln \prod_{(h(x), z) \in S} p(z|h(x)) = -\sum_{(h(x), z) \in S} \ln p(z|h(x))$ 评估目标模型的误差。

[0081] 其中, $L(S)$ 为误差, x 为语音信息, z 为文本标签, $p(z|h(x))$ 为预测文本与文本标签的相似度, S 为多个训练语音样本。预测文本是指语音信息输入至目标模型后,由目标模型根据神经元权值计算输出的文本信息。

[0082] CTC损失函数用来估量语音识别模型输出的预测文本与真实的文本标签的不一致程度,其优点是不要输入数据与输出数据强制对齐。与输入特征和目标标签之间的帧级对准的交叉熵准则不同,CTC损失函数能够自动学习语音数据和标签序列(比如,音素或者字符等)之间的对齐,这消除了对数据进行强制对齐的需要,并且输入数据与标签的长度不一定相同。以CTC损失函数评估语音识别模型的预测文本,无需考虑文本标签中的发音音素与训练语音信息的序列之间精准的映射关系,只需要输入序列和输出序列即可训练语音识别模型,节省了训练语音样本集的制作成本。

[0083] 104、调整目标模型的神经元的权值,直至误差小于阈值,将误差小于阈值的神经元权值设为理想权值。

[0084] 根据CTC损失函数计算出相应的训练语音样本集的误差,通过梯度下降算法在语音识别模型中的反向传播误差,从而更新语音识别模型中的权重与阈值等目标参数,不断提高语音识别模型语音识别的准确率,直至达到收敛要求。

[0085] 105、将目标模型以及理想权值部署至客户端。

[0086] 相较于现有技术,本申请通过旁路通道将输入信息 x 直接绕道传至隐含层的输出,旁路通道没有权值,保护了输入信息 x 的完整性,使得神经网络训练的更深,整个神经网络只需要训练输入、输出差别的部分,即传递输入信息 x 后,每个残差模块只学习残差 $F(x)$,简化训练的目标和难度,且神经网络稳定易于训练,随着神经网络深度的增加,语音识别模型的性能也逐渐变好,并且以CTC损失函数评估语音识别模型的预测文本,无需考虑文本标签中的发音音素与训练语音信息的序列之间精准的映射关系,只需要输入序列和输出序列即可训练语音识别模型,节省了训练语音样本集的制作成本。此外,采用三角带通滤波器对训练语音信息的频谱进行平滑,消除训练语音信息中的谐波,突显原始声音的共振峰,避免语音信息中的音调对语音识别模型预测文本的影响,并降低语音识别模型识别过程中的对语音信息的运算量。

[0087] 一些实施方式中,将多个语音样本输入至语音识别模型之前,方法还包括:

[0088] 根据预设的分帧参数分帧处理训练语音信息,得到训练语音信息对应的语句,预设分帧参数包括帧时长、帧数和前后帧重复时长;

[0089] 根据预设的二维参数和滤波器组特征提取算法转化语句,得到二维语音信息。

[0090] 一些实施方式中,根据预设的分帧参数分帧处理训练语音信息,包括:

[0091] 对二维语音信息进行离散傅里叶变换,以得到二维语音信息对应的线性频谱 $X(k)$;;

[0092] 通过预设的带通滤波器对线性频谱滤波,以得到目标线性频谱,当带通滤波器的

中心频率为 $f(m)$ 时,则带通滤波器的传递函数为:

$$[0093] \quad H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad f(m) \text{的表达式为:}$$

$$[0094] \quad f(m) = \left(\frac{N}{f_s} \right) F_{mel}^{-1} \left(F_{mel}(f_l) \right) + m \frac{F_{mel}(f_h) - F_{mel}(f_l)}{M+1},$$

[0095] 带通滤波器包括多个具有三角形滤波特性的带通滤波器, f_l 为带通滤波器频率范围的最低频率, f_h 为带通滤波器频率范围的最高频率, N 为DFT时的长度, f_s 为带通滤波器的采样频率, F_{mel} 函数为 $F_{mel} = 1125 \ln(1+f/700)$,

[0096] F_{mel} 的逆函数为: $F_{mel}^{-1}(b) = 700(e^{b/1125} - 1)$, b 为整数;

[0097] 根据 $s(m) = \ln(\sum_{k=0}^{N-1} |X(k)|^2 H_m(k))$, $0 \leq m \leq M$ 计算目标线性频谱对应的对数能量,得到语谱图, $X(k)$ 为线性频谱。

[0098] 上述实施方式中,人对声音声压的反应呈对数关系,人对高声压的细微变化敏感度不如低声压。此外,使用对数可以降低提取的特征对输入声音能量变化的敏感度,因为声音与麦克风之间的距离是变化的,因而麦克风采集到的声音能量也是变化的。语谱图是一种声音能量时频分布的可视化表达方式,有效的利用了时频两域之间的相关性,通过语谱图分析获得的特征矢量序列对于声学特征的提取的效果更好,输入到语音识别模型中,使后续的运算准确性更高。并且采用三角带通滤波器对训练语音信息的频谱进行平滑,消除训练语音信息中的谐波,突显原始声音的共振峰。因此训练语音信息中一段声音的音调或音高,不会反应在声学特征内,也就是说,语音识别模型不会受到语音信息中的音调不同而对预测文本有所影响;并且降低了语音识别模型识别过程中的对语音信息的运算量。

[0099] 一些实施方式中,全连接层包括分类函数。分类函数是指 $\delta(z)_j = \frac{\exp(z_j)}{\sum_{k=1}^K \exp(z_k)}$ 。 j 为

自然数,分类函数将卷积残差层输出的 K 维的语音频域信号向量 z 压缩到另一个 K 维实向量 $\delta(z)_j$,使得每一个元素的范围都在 $(0, 1)$ 之间,并且所有元素的和为1。

[0100] 一些实施方式中,残差模块的输入为 x ,输出残差模块的输出为 y ,则残差模块的数学表达式为:

[0101] $y = F(x, w_i) + w_s x$ 。 $F(x, w_i)$ 为独立卷积层的输出, w_s 为残差模块的权值。

[0102] 上述实施方式中,本实施例中的语音识别模型为若干个顺次连接的隐藏层增加了旁路通道,以解决传统的神经网络随着网络层数的增加,训练准确率越来越低的问题。语音识别模型的卷积残差层有很多旁路通道,旁路通道作为隐藏层的支线,实现隐藏层之间的跨层连接,即将隐藏层的输入直接连到下一级层,使得下一级层可以直接学习残差。

[0103] 具体地,在一个残差模块中,跨层连接一般只跨越2至3个隐藏层,但不排斥跨越更多的隐藏层。仅跨越1个隐藏层的情况意义不大,实验效果不理想。

[0104] 假定残差模块的输入为 x ,期望输出是 $H(x)$,即 $H(x)$ 是期望的复杂潜在映射,但通常 $H(x)$ 的学习难度很大;如果直接把输入 x 传到输出作为初始结果,那么此时残差模块需要学习的目标就是 $F(x) = H(x) - x$ 。于是,相对于传统的神经网络,本实施例中的语音识别模型相当于将学习目标改变了,不再是学习一个完整的输出,而是学习最优解 $H(x)$ 和全等映射 x 的差值,即残差: $F(x) = H(x) - x$ 。从整体功能上看,如果用 $\{w_i\}$ 表示残差模块的所有权值,那么残差模块实际上计算的输出结果为: $y = F(x, \{w_i\}) + x$,以跨越2个隐藏层为例,在忽略偏置的情况下, $F(x, \{w_i\}) = w_2 \delta(w_1 x) = w_2 \text{ReLU}(w_1 x)$,其中, $\text{ReLU}()$ 为残差模块的激活函数。

[0105] 可以理解的是, $F(x, \{w_i\})$ 与 x 需要具有相同的维数。如果它们的维数不相同,则可以引入一个额外的权值矩阵 w_s 对 x 进行线性投影,使得 $F(x, \{w_i\})$ 与 x 的维数相同,相应地,残差模块的计算结果为: $y = F(x, \{w_i\}) + w_s x$

[0106] 一些实施方式中, $F(x, w_i)$ 的采用ReLU函数作为独立卷积层的激活函数,ReLU函数的数学表达式为 $\text{ReLU}(x) = \max(0, x)$ 。

[0107] 上述实施方式中,通过上述公式可以训练神经网络。

[0108] 一些实施方式中,调整目标模型的神经元的权值,包括:

[0109] 通过随机梯度下降法调整神经元的权值。

[0110] 上述实施方式中,采用随机梯度下降算法能有效避免冗余计算,消耗时间更短。当然本领域技术人员还可以采用其它算法。

[0111] 如图2所示的一种构建语音识别模型的装置20的结构示意图,其可应用于构建语音识别模型。本申请实施例中的构建语音识别模型的装置能够实现对应于上述图1所对应的实施例中所执行的构建语音识别模型的方法的步骤。构建语音识别模型的装置20实现的功能可以通过硬件实现,也可以通过硬件执行相应的软件实现。硬件或软件包括一个或多个与上述功能相对应的模块,所述模块可以是软件和/或硬件。所述构建语音识别模型的装置可包括输入输出模块201和处理模块202,所述处理模块202和输入输出模块201的功能实现可参考图1所对应的实施例中所执行的操作,此处不作赘述。输入输出模块201可用于控制所述输入输出模块201的输入、输出以及获取操作。

[0112] 一些实施方式中,所述输入输出模块201可用于用于获取多个训练语音样本,所述训练语音样本包括语音信息以及与语音信息对应的文本标签;

[0113] 所述处理模块202可用于用于通过独立卷积层、卷积残差层、全连接层以及输出层构建语音识别模型,所述卷积残差层包括多个顺次连接的残差堆叠层,所述残差堆叠层包含多个顺次连接的残差模块,所述残差模块包含多个顺次连接的隐藏层以及旁路于多个顺次连接的权值层的旁路通道;通过所述输入输出模块将多个所述语音样本依次输入至所述语音识别模型,将所述语音信息及将所述语音信息对应的文本标签分别作为所述语音识别模型的输入以及输出,通过所述输入以及所述输出不断训练所述语音识别模型的神经元权值,直至所述语音样本均已输入至所述语音识别模型,结束对所述语音识别模型的训练,所述训练结束后,将带有训练好神经元权值的所述语音识别模型作为目标模型;通过 $L(S) = -\ln \prod_{(h(x), z) \in S} p(z|h(x)) = -\sum_{(h(x), z) \in S} \ln p(z|h(x))$ 评估所述目标模型的误差;其中, $L(S)$ 为所述误差, x 为所述语音信息, z 为所述文本标签, $p(z|h(x))$ 为所述预测文本与所述文本标签的相似度, S 为所述多个训练语音样本,所述预测文本是指所述语音信息输入至所述目标模型后,由所述目标模型根据神经元权值计算输出的文本信息;调整所述目标模型的神

经元的权值,直至所述误差小于阈值,将所述误差小于阈值的神经元权值设为理想权值。将所述目标模型以及所述理想权值部署至客户端。

[0114] 一些实施方式中,所述处理模块202还用于:

[0115] 根据预设的分帧参数分帧处理所述训练语音信息,得到所述训练语音信息对应的语句,所述预设分帧参数包括帧时长、帧数和前后帧重复时长;

[0116] 根据预设的二维参数和滤波器组特征提取算法转化所述语句,得到二维语音信息。

[0117] 一些实施方式中,所述处理模块202还用于:

[0118] 对所述二维语音信息进行离散傅里叶变换,以得到所述二维语音信息对应的线性频谱 $X(k)$;

[0119] 通过预设的带通滤波器对所述线性频谱滤波,以得到目标线性频谱,当所述带通滤波器的中心频率为 $f(m)$ 时,则所述带通滤波器的传递函数为:

$$[0120] \quad H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad \text{所述 } f(m) \text{ 的表达式为:}$$

$$[0121] \quad f(m) = \left(\frac{N}{f_s} \right) F_{mel}^{-1} \left(F_{mel}(f_l) + m \frac{F_{mel}(f_h) - F_{mel}(f_l)}{M+1} \right),$$

[0122] 所述带通滤波器包括多个具有三角形滤波特性的带通滤波器,所述 f_l 为所述带通滤波器频率范围的最低频率,所述 f_h 为所述带通滤波器频率范围的最高频率,所述 N 为DFT时的长度,所述 f_s 为所述带通滤波器的采样频率,所述 F_{mel} 函数为 $F_{mel} = 1125 \ln(1+f/700)$,所述 F_{mel} 的逆函数为: $F_{mel}^{-1}(b) = 700(e^{b/1125} - 1)$, b 为整数;

[0123] 根据 $s(m) = \ln \left(\sum_{k=0}^{N-1} |X(k)^2 H_m(k)| \right)$, $0 \leq m \leq M$ 计算所述目标线性频谱对应的对数能量,得到语谱图,所述 $X(k)$ 为所述线性频谱;

[0124] 一些实施方式中,所述全连接层包括分类函数,所述分类函数是指

$$\delta(z)_j = \frac{\exp(z_j)}{\sum_{k=1}^K \exp(z_k)}, \text{ 所述 } j \text{ 为自然数, 所述分类函数将卷积残差层输出的 } K \text{ 维的语音频域信号向量 } z \text{ 压缩到另一个 } K \text{ 维实向量, 使得每一个元素的范围都在 } (0, 1) \text{ 之间, 并且所有元素的和为 } 1.$$

[0125] 一些实施方式中,所述残差模块的输入为 x ,所述输出残差模块的输出为 y ,则所述残差模块的数学表达式为: $y = F(x, w_i) + w_s x$,所述 $F(x, w_i)$ 为所述独立卷积层的输出,所述 w_s 为所述残差模块的权值。

[0126] 一些实施方式中,其特征在于,所述 $F(x, w_i)$ 的采用ReLU函数作为所述独立卷积层的激活函数,所述ReLU函数的数学表达式为 $\text{ReLU}(x) = \max(0, x)$ 。

[0127] 一些实施方式中,所述调整所述目标模型的神经元的权值,包括:

[0128] 通过随机梯度下降法调整所述神经元的权值。

[0129] 上面从模块化功能实体的角度分别介绍了本申请实施例中的创建装置,以下从硬件角度介绍一种构建语音识别模型的设备,如图3所示,其包括:处理器、存储器、输入输出单元(也可以是收发器,图3中未标识出)以及存储在所述存储器中并可在所述处理器上运行的计算机程序。例如,该计算机程序可以为图1所对应的实施例中构建语音识别模型的方法对应的程序。例如,当构建语音识别模型的设备实现如图2所示的构建语音识别模型的装置20的功能时,所述处理器执行所述计算机程序时实现上述图2所对应的实施例中由构建语音识别模型的装置20执行的构建语音识别模型的方法中的各步骤。或者,所述处理器执行所述计算机程序时实现上述图2所对应的实施例的构建语音识别模型的装置20中各模块的功能。又例如,该计算机程序可以为图1所对应的实施例中构建语音识别模型的方法对应的程序。

[0130] 所称处理器可以是中央处理单元(central processing unit,CPU),还可以是其他通用处理器、数字信号处理器(digital signal processor,DSP)、专用集成电路(application specific integrated circuit,ASIC)、现成可编程门阵列(field-programmable gate array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等,所述处理器是所述计算机装置的控制中心,利用各种接口和线路连接整个计算机装置的各个部分。

[0131] 所述存储器可用于存储所述计算机程序和/或模块,所述处理器通过运行或执行存储在所述存储器内的计算机程序和/或模块,以及调用存储在存储器内的数据,实现所述计算机装置的各种功能。所述存储器可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等;存储数据区可存储根据手机的使用所创建的数据(比如音频数据、视频数据等)等。此外,存储器可以包括高速随机存取存储器,还可以包括非易失性存储器,例如硬盘、内存、插接式硬盘,智能存储卡(smart media card,SMC),安全数字(secure digital,SD)卡,闪存卡(Flash Card)、至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

[0132] 所述输入输出单元也可以用接收器和发送器代替,可以为相同或者不同的物理实体。为相同的物理实体时,可以统称为输入输出单元。该输入输出可以为收发器。

[0133] 所述存储器可以集成在所述处理器中,也可以与所述处理器分开设置。

[0134] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM)中,包括若干指令用以使得一台终端(可以是手机,计算机,服务器或者网络设备)执行本申请各个实施例所述的方法。

[0135] 上面结合附图对本申请的实施例进行了描述,但是本申请并不局限于上述的具体实施方式,上述的具体实施方式仅仅是示意性的,而不是限制性的,本领域的普通技术人员在本申请的启示下,在不脱离本申请宗旨和权利要求所保护的范围情况下,还可做出很多形式,凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换,或直接或间接

运用在其他相关的技术领域,这些均属于本申请的保护之内。

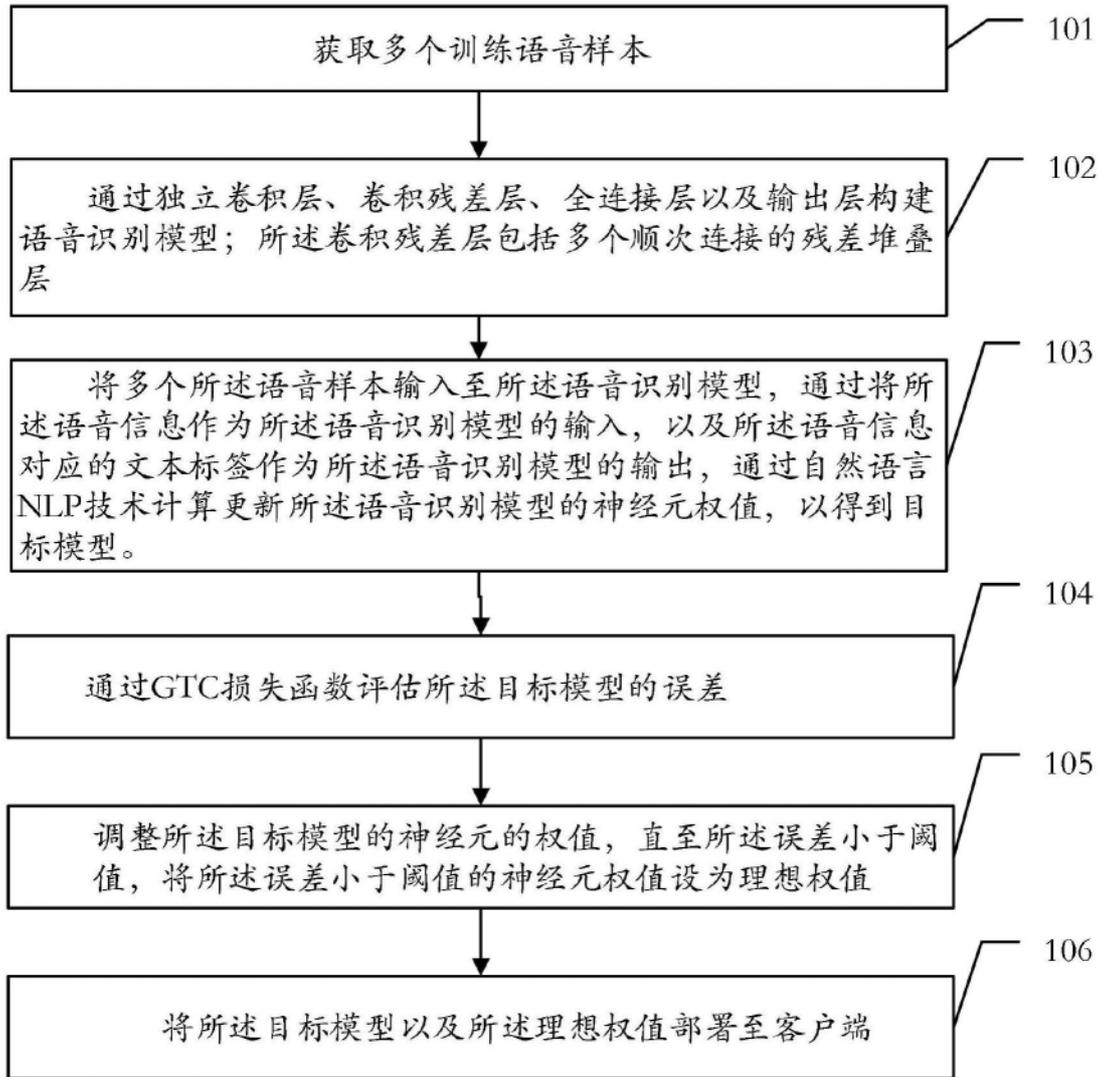


图1

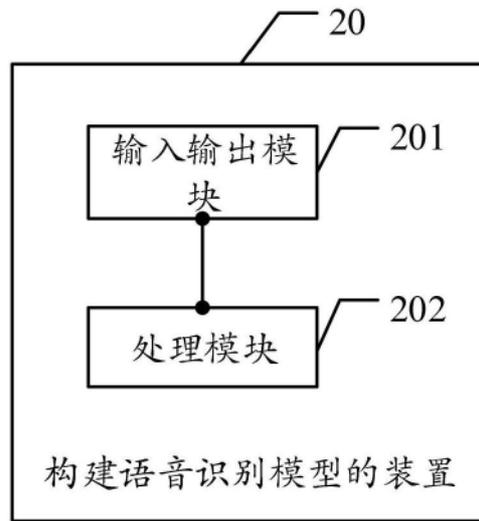


图2

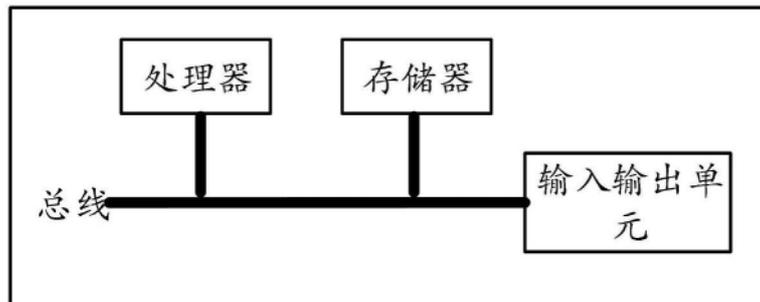


图3