(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109410917 A
(43)申请公布日 2019.03.01

(21)申请号 201811123791.1

(22)申请日 2018.09.26

(71)申请人 河海大学常州校区
地址 213022 江苏省常州市晋陵北路200号

(72)发明人 徐宁 倪亚南 刘小峰 潘安顺 刘妍妍

(74)专利代理机构 南京纵横知识产权代理有限公司 32224
代理人 董建林

(51)Int.Cl.
G10L 15/02(2006.01)
G10L 15/06(2013.01)
G10L 15/08(2006.01)
G06N 3/04(2006.01)

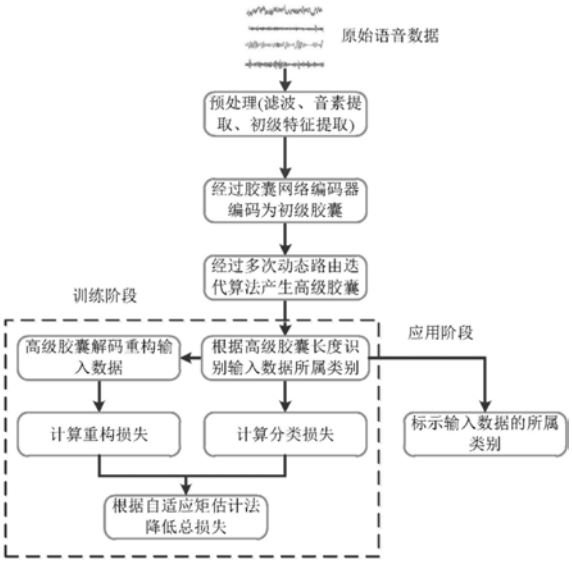
权利要求书3页 说明书8页 附图2页

(54)发明名称

基于改进型胶囊网络的语音数据分类方法

(57)摘要

本发明公开了一种基于改进型胶囊网络的语音数据分类方法，首先在训练阶段，构造胶囊网络的编码器，对初始语音音素数据进行编码得到初级胶囊；构造胶囊网络的动态路由结构，将初级胶囊中的信息传递给高级胶囊；以每个高级胶囊的长度的softmax激活值表征初始语音音素数据属于对应类别的概率；然后构造胶囊网络的解码器，将真实音素符号对应的高级胶囊进行解码重构；基于总损失函数对胶囊网络的参数进行优化；在测试阶段，将初始语音音素数据输入胶囊网络的编码器中，判断待测试数据的所属类别。本发明实现对语音音素的时序信号所对应的音速符号的精准识别，解决按照理论知识直接提取的特征准确度低、语音数据分类效果差以及过拟合的技术问题。



CN 109410917 A

1.基于改进型胶囊网络的语音数据分类方法,其特征在于,包括以下步骤:

训练阶段:

1)构造胶囊网络的编码器,具体为,

11)利用神经网络的前向传播算法对初始语音音素数据进行编码,得到初级胶囊;

12)构造胶囊网络的动态路由结构,将初级胶囊中的信息传递给高级胶囊;

13)以每个高级胶囊的长度的softmax激活值表征初始语音音素数据属于对应类别的概率;

2)构造胶囊网络的解码器,将真实音素符号对应的高级胶囊进行解码重构;

3)基于预设的损失函数得到总损失,对胶囊网络的参数进行优化,最小化总损失;

测试阶段:

4)将待测试的初始语音音素数据输入胶囊网络的编码器中,根据所有高级胶囊长度的softmax激活值判断待测试的初始语音音素数据的所属类别。

2.根据权利要求1所述的基于改进型胶囊网络的语音数据分类方法,其特征在于,所述初始语音音素数据是通过原始语音数据经过预处理得到,具体包括以下步骤:

A.对原始语音进行带通滤波平滑处理,去除毛刺噪声点;

B.对滤波后的语音信号进行分帧处理;

C.对分帧后的每一帧语音信号进行特征提取,选取w个特征作为每一帧的w维特征向量;

D.对每一帧的w维特征向量进行归一化处理得到初始语音音素数据,即将每一帧的w维特征向量按维度除以一个归一化因子,所述归一化因子是一个w维向量l_norm。

3.根据权利要求1所述的基于改进型胶囊网络的语音数据分类方法,其特征在于,所述步骤11)中所述编码方法具体为全连接网络的编码、二维卷积网络的编码或者混合型编码。

4.根据权利要求3所述的基于改进型胶囊网络的语音数据分类方法,其特征在于,所述步骤11)利用混合型编码得到初级胶囊的具体步骤为:

111)初始语音音素数据经过一个全连接层连接到具有m_1个单元的隐藏层h_1,经过激活函数sigmoid作用得到隐藏层h_1的输出数据z_1;

112)将隐藏层h_1的输出数据z_1重塑为一个三维张量input1,即m_2个b_1*b_1大小的特征图谱,通过卷积核大小为k_1*k_1的二维卷积层连接到具有m_3个b_2*b_2大小的特征图谱的隐藏层h_2,经过激活函数sigmoid作用得到隐藏层h_2的输出z_2;

113)将隐藏层h_2的输出数据z_2重塑成n_1*pri_len大小的二维张量,代表n_1个初级胶囊,张量矩阵的每一行代表一个初级胶囊pri_cap,每个初级胶囊的维度是pri_len;

114)将每个初级胶囊按照如下公式进行squash归一化处理。

$$\mathrm{squash}(pri\_cap)=\frac{\left\|pri\_cap\right\|^2}{1+\left\|pri\_cap\right\|^2}\times\frac{pri\_cap}{\left\|pri\_cap\right\|}$$

5.根据权利要求1所述的基于改进型胶囊网络的语音数据分类方法,其特征在于,所述步骤12)构造胶囊网络的动态路由结构,将初级胶囊的信息传递给高级胶囊,高级胶囊的个数即为预定的分类类别数目,动态路由结构采用路由迭代,迭代次数为预设值,具体步骤为:

121)将n_1个维度为pri_len的初级胶囊复制n_2份得到一个n_1*n_2*pri_len*1的张量tile_pri_cap,n_2为预设的高级胶囊数目,基于张量的矩阵乘法得到中间张量p,具体公式如下:

p=W·tile_pri_cap

其中W是一个形如n_1*n_2*w_n*pri_len的权值张量,w_n为高级胶囊维度,·代表张量的矩阵乘法,即执行n_1*n_2次w_n*pri_len与pri_len*1的矩阵相乘,p是形如n_1*n_2*w_n*1的动态路由的中间张量,其中W是可迭代更新的参数;

122)进行动态路由的迭代,根据softmax规则归一化张量B的每一行得到激活后的耦合系数张量C,耦合系数张量C中的C_{ij}代表中间张量p中第i个初级胶囊的信息分配到第j个高级胶囊的比例,初次迭代时,张量B为形如n_1*n_2的全零张量;

将激活后的耦合系数张量C根据张量乘法乘以中间张量p,具体公式如下:

S=multiply(C,p)

其中multiply代表p中的n_1*n_2个w_n维向量按对应位置乘以C中的n_1*n_2个耦合系数,S是形如n_1*n_2*w_n*1的媒介张量,代表经过耦合系数进行初级胶囊和高级胶囊之间信息传递后的中间信息媒介;

123)将媒介张量S按照第一维度进行求和操作并且保留维度,将第一个维度轴上的n_1个数字相加,得到n_2个维度是w_n的高级胶囊,再使用squash函数对高级胶囊进行归一化处理,得到形如1*n_2*w_n的张量,如果本次迭代为最后一次动态路由的迭代,则输出归一化后的高级胶囊pho_cap,进入步骤13);如果本次迭代不是最后一次迭代,则进入步骤124);

124)将归一化后的高级胶囊pho_cap按照第一个维度复制n_1份,n_1为初级胶囊个数,得到形如n_1*n_2*w_n的张量v_j,将张量v_j扩增维度得到形如n_1*n_2*1*w_n的张量,按张量的矩阵乘法计算张量v_j与中间张量p的乘积得到形如n_1*n_2*1的张量胶囊的相似性度量矩阵,将相似性度量矩阵张量按元素对应加到张量B上,至此完成一次动态路由的迭代过程,进入步骤122)。

6.根据权利要求1所述的基于改进型胶囊网络的语音数据分类方法,其特征在于,所述步骤2)构造胶囊网络的解码器,将真实音素符号对应的高级胶囊进行解码重构具体步骤为:

21)将非真实音素符号对应的高级胶囊中的数据置零,并将n_2个维度w_n为高级胶囊的数据重塑成n_2*w_n维度的向量v_pho_cap;

22)将向量v_pho_cap经过包括隐藏层de_fc1和隐藏层de_fc2的全连接层,以全连接的方式连接到输出维度为w的解码器输出层,w为初始语音音素数据的维度,解码出的重构数据与初始语音音素数据具有相同的数据格式。

7.根据权利要求1所述的基于改进型胶囊网络的语音数据分类方法,其特征在于,所述步骤3)基于预设的损失函数得到总损失,对胶囊网络的参数进行优化,最小化总损失,具体步骤如下:

31)采用解码器输出的重构数据与初始语音音素数据之间距离的平方差作为重构损失L_r;

32)根据所有高级胶囊的长度得到分类损失L_c,分类损失表示高级胶囊的长度与期望值阈值的差距,计算公式如下:

$$L\_c=\sum_k T_k \max\left(0, m^+ - \left\|pho\_cap_k\right\|\right)^2 + \lambda\left(1-T_k\right)\max\left(0, \left\|pho\_cap_k\right\| - m^-\right)^2$$

其中k是预测音素符号，T_k是一个分类标签指示函数，当k指向真实音素符号时，T_k为1，||pho_cap_k||表示预测音素符号k对应的归一化后的高级胶囊的长度，m^+和m^-分别为归一化后的单个高级胶囊长度的上下限值，max函数表示取两个数值中较大的一个，即当真实音素符号对应的高级胶囊长度超过上限，则该高级胶囊的分类损失忽略，非真实音素符号对应的高级胶囊长度低于下限时，则该高级胶囊的分类损失忽略，λ为非真实音素符号对应的分类损失之和的缩放系数；

33)总损失函数L_t＝L_c+η•L_r，其中η是用于调整分类损失和重构损失在总损失中的比重的比例系数，采用自适应矩估计法对总损失函数进行优化，即根据链式求导以及反向传播法则求出总损失函数对胶囊网络中每一个可更新参数的导数，进而最小化总损失，优化胶囊网络。

8.根据权利要求7所述的基于改进型胶囊网络的语音数据分类方法，其特征在于，所述步骤33)中最小化总损失具体为：采用小批量的梯度下降法，在训练时每次前向计算的总损失为小批量数据中所有输入数据对应的总损失的平均值；使用自适应矩估计法对胶囊网络中所有可更新参数进行求导，并使用梯度下降法对每个小批量训练数据更新胶囊网络中的可更新参数。

基于改进型胶囊网络的语音数据分类方法

技术领域

[0001] 本发明属于分类处理及深度学习技术领域,具体涉及一种基于改进型胶囊网络的语音数据分类方法。

背景技术

[0002] 语音数据是现代信息数据的重要处理内容,每一帧语音数据都可以用特征参数来描绘,比如共振峰有关参数,即一帧语音数据的共振峰频率(第一维)、带宽(第二维)、能量频谱倾斜(第三维)等,以上是基于研究人员经验积累、按照理论知识直接提取出来的多维特征。然而这样的工作计算量非常大,并且需要大量的尝试以及创新。近几年崛起的深度学习方法集特征提取和特征的分类于一体,具有非常强大的特征自组织以及特征抽象能力,能够帮助研究人员减轻在语音数据特征的设计上投入的时间和精力。卷积神经网络目前已经在图像的分类识别方面取得了巨大的成就,但是由于语音数据和图像数据存在一定的差异,卷积神经网络并不适合直接处理语音数据。

发明内容

[0003] 本发明的目的在于,提出一种基于改进型胶囊网络的语音数据分类方法,实现对语音音素的时序信号所对应的音速符号的精准识别,解决现有技术中按照理论知识直接提取的语音特征准确度低、语音数据分类效果差以及过拟合的技术问题。

[0004] 本发明采用如下技术方案,一种基于改进型胶囊网络的语音数据分类方法,具体包括:

[0005] 训练阶段:

[0006] 1)构造胶囊网络的编码器,具体为,

[0007] 11)利用神经网络的前向传播算法对初始语音音素数据进行编码,得到初级胶囊;

[0008] 12)构造胶囊网络的动态路由结构,将初级胶囊中的信息传递给高级胶囊;

[0009] 13)以每个高级胶囊的长度的softmax激活值表征初始语音音素数据属于对应类别的概率;

[0010] 2)构造胶囊网络的解码器,将真实音素符号对应的高级胶囊进行解码重构;

[0011] 3)基于预设的损失函数得到总损失,对胶囊网络的参数进行优化,最小化总损失;

[0012] 测试阶段:

[0013] 4)将待测试的初始语音音素数据输入胶囊网络的编码器中,根据所有高级胶囊长度的softmax激活值判断待测试的初始语音音素数据的所属类别。

[0014] 优选地,所述初始语音音素数据是通过原始语音数据经过预处理得到,具体包括以下步骤:

[0015] A.对原始语音进行带通滤波平滑处理,去除毛刺噪声点;

[0016] B.对滤波后的语音信号进行分帧处理;

[0017] C.对分帧后的每一帧语音信号进行特征提取,选取w个特征作为每一帧的w维特征

向量；

[0018] D.对每一帧的w维特征向量进行归一化处理得到初始语音音素数据，即将每一帧的w维特征向量按维度除以一个归一化因子，所述归一化因子是一个w维向量l_norm。

[0019] 优选地，所述步骤11)中所述编码方法具体为全连接网络的编码、二维卷积网络的编码或者混合型编码。

[0020] 优选地，所述步骤11)利用混合型编码得到初级胶囊的具体步骤为：

[0021] 111)初始语音音素数据经过一个全连接层连接到具有m_1个单元的隐藏层h_1，经过激活函数sigmoid作用得到隐藏层h_1的输出数据z_1；

[0022] 112)将隐藏层h_1的输出数据z_1重塑为一个三维张量input1，即m_2个b_1*b_1大小的特征图谱，通过卷积核大小为k_1*k_1的二维卷积层连接到具有m_3个b_2*b_2大小的特征图谱的隐藏层h_2，经过激活函数sigmoid作用得到隐藏层h_2的输出z_2；

[0023] 113)将隐藏层h_2的输出数据z_2重塑成n_1*pri_len大小的二维张量，代表n_1个初级胶囊，张量矩阵的每一行代表一个初级胶囊pri_cap，每个初级胶囊的维度是pri_len；

[0024] 114)将每个初级胶囊按照如下公式进行squash归一化处理。

[0025] $$\mathrm{squash}(pri\_cap)=\frac{\|pri\_cap\|^2}{1+\|pri\_cap\|^2}\times\frac{pri\_cap}{\|pri\_cap\|}$$

[0026] 优选地，所述步骤12)构造胶囊网络的动态路由结构，将初级胶囊的信息传递给高级胶囊，高级胶囊的个数即为预定的分类类别数目，动态路由结构采用路由迭代，迭代次数为预设值，具体步骤为：

[0027] 121)将n_1个维度为pri_len的初级胶囊复制n_2份得到一个$n_1*n_2*pri_len*1$的张量tile_pri_cap，n_2为预设的高级胶囊数目，基于张量的矩阵乘法得到中间张量p，具体公式如下：

[0028] $p=W\cdot tile\_pri\_cap$

[0029] 其中W是一个形如$n_1*n_2*w_n*pri_len$的权值张量，w_n为高级胶囊维度，·代表张量的矩阵乘法，即执行n_1*n_2次w_n*pri_len与pri_len*1的矩阵相乘，p是形如$n_1*n_2*w_n*1$的动态路由的中间张量，其中W是可迭代更新的参数；

[0030] 122)进行动态路由的迭代，根据softmax规则归一化张量B的每一行得到激活后的耦合系数张量C，耦合系数张量C中的C_{ij}代表中间张量p中第i个初级胶囊的信息分配到第j个高级胶囊的比例，初次迭代时，张量B为形如n_1*n_2的全零张量；

[0031] 将激活后的耦合系数张量C根据张量乘法乘以中间张量p，具体公式如下：

[0032] $S=multiply(C,p)$

[0033] 其中multiply代表p中的n_1*n_2个w_n维向量按对应位置乘以C中的n_1*n_2个耦合系数，S是形如$n_1*n_2*w_n*1$的媒介张量，代表经过耦合系数进行初级胶囊和高级胶囊之间信息传递后的中间信息媒介；

[0034] 123)将媒介张量S按照第一维度进行求和操作并且保留维度，将第一个维度轴上的n_1个数字相加，得到n_2个维度是w_n的高级胶囊，再使用squash函数对高级胶囊进行归一化处理，得到形如$1*n_2*w_n$的张量，如果本次迭代为最后一次动态路由的迭代，则输出归一化后的高级胶囊pho_cap，进入步骤13)；如果本次迭代不是最后一次迭代，则进入步骤124)；

[0035] 124)将归一化后的高级胶囊pho_cap按照第一个维度复制n_1份，n_1为初级胶囊个数，得到形如$n_1*n_2*w_n$的张量v_j，将张量v_j扩增维度得到形如$n_1*n_2*1*w_n$的张量，按张量的矩阵乘法计算张量v_j与中间张量p的乘积得到形如n_1*n_2*1的张量胶囊的相似性度量矩阵，将相似性度量矩阵张量按元素对应加到张量B上，至此完成一次动态路由的迭代过程，进入步骤122)；

[0036] 优选地，所述步骤2)构造胶囊网络的解码器，将真实音素符号对应的高级胶囊进行解码重构具体步骤为：

[0037] 21)将非真实音素符号对应的高级胶囊中的数据置零，并将n_2个维度w_n为高级胶囊的数据重塑成n_2*w_n维度的向量v_pho_cap；

[0038] 22)将向量v_pho_cap经过包括隐藏层de_fc1和隐藏层de_fc2的全连接层，以全连接的方式连接到输出维度为w的解码器输出层，w为初始语音音素数据的维度，解码出的重构数据与初始语音音素数据具有相同的数据格式。

[0039] 优选地，所述步骤3)基于预设的损失函数得到总损失，对胶囊网络的参数进行优化，最小化总损失，具体步骤如下：

[0040] 31)采用解码器输出的重构数据与初始语音音素数据之间距离的平方差作为重构损失L_r；

[0041] 32)根据所有高级胶囊的长度得到分类损失L_c，分类损失表示高级胶囊的长度与期望值阈值的差距，计算公式如下：

[0042] $$L\_c=\sum_k T_k \max\left(0, m^+-\left\|pho\_cap_k\right\|\right)^2+\lambda\left(1-T_k\right)\max\left(0, \left\|pho\_cap_k\right\|-m^-\right)^2$$

[0043] 其中k是预测音素符号，T_k是一个分类标签指示函数，当k指向真实音素符号时，T_k为1，||pho_cap_k||表示预测音素符号k对应的归一化后的高级胶囊的长度，m^+和m^-分别为归一化后的单个高级胶囊长度的上下限值，max函数表示取两个数值中较大的一个，即当真实音素符号对应的高级胶囊长度超过上限，则该高级胶囊的分类损失忽略，非真实音素符号对应的高级胶囊长度低于下限时，则该高级胶囊的分类损失忽略，λ为非真实音素符号对应的分类损失之和的缩放系数；

[0044] 33)总损失函数L_t＝L_c+η·L_r，其中η是用于调整分类损失和重构损失在总损失中的比重的比例系数，采用自适应矩估计法对总损失函数进行优化，即根据链式求导以及反向传播法则求出总损失函数对胶囊网络中每一个可更新参数的导数，进而最小化总损失，优化胶囊网络。

[0045] 优选地，所述步骤33)中最小化总损失具体为：采用小批量的梯度下降法，在训练时每次前向计算的总损失为小批量数据中所有输入数据对应的总损失的平均值；使用自适应矩估计法对胶囊网络中所有可更新参数进行求导，并使用梯度下降法对每个小批量训练数据更新胶囊网络中的可更新参数

[0046] 发明所达到的有益效果：本发明是一种基于改进型胶囊网络的语音数据分类方法，实现对语音音素的时序信号所对应的音速符号的精准识别，解决现有技术中按照理论知识直接提取的特征准确度低、语音数据分类效果差以及过拟合的技术问题。本发明提取出频域系数等多个维度的特征表征的语音音素数据，颗粒度精细到每一帧，可以对每一帧的音素进行识别，具有较高的实时性；语音数据输入胶囊网络后，经过卷积层以及动态路

由,可以得到表征所要求的类别的高级胶囊向量,根据各个高级胶囊的长度的softmax激活值可以判断出输入数据属于各个类别的概率;在训练阶段根据高级胶囊解码重构出输入数据,大大增强了所提取的高级胶囊对输入数据表征能力的置信度。

附图说明

[0047] 图1为本发明一种实施例的基于改进型胶囊网络的语音数据分类方法流程图;

[0048] 图2为本发明一种实施例中胶囊网络整体架构中的编码器架构示意图;

[0049] 图3为本发明一种实施例中胶囊网络整体架构中的解码器架构示意图;

[0050] 图4为本发明一种实施例中的squash函数示意图。

具体实施方式

[0051] 下面根据附图并结合实施例对本发明的技术方案作进一步阐述,以下实施例只是描述性的,不是限定性的,不能以此限定本发明的保护范围。

[0052] 图1为本发明一种实施例的基于改进型胶囊网络的语音数据分类方法流程图。

[0053] 基于改进型胶囊网络的语音数据分类方法,具体步骤如下:

[0054] 训练阶段:

[0055] 1)构造胶囊网络的编码器,如图2所示,具体为,

[0056] 11)利用神经网络的前向传播算法对初始语音音素数据进行编码,得到初级胶囊;具体的公式为:

[0057] pri_cap=forward(input0)

[0058] 其中,pri_cap表示编码后的初级胶囊向量,forward表示编码过程,input0表示初始语音音素数据,即编码器部分的输入数据;

[0059] 12)构造胶囊网络的动态路由结构,将初级胶囊中的信息传递给高级胶囊;

[0060] 13)以每个高级胶囊的长度的softmax激活值表征初始语音音素数据属于对应类别的概率;

[0061] 2)构造胶囊网络的解码器,将真实音素符号对应的高级胶囊进行解码重构;

[0062] 3)基于预设的损失函数得到总损失,对胶囊网络的参数进行优化,最小化总损失;

[0063] 测试阶段:

[0064] 4)将待测试的初始语音音素数据输入胶囊网络的编码器中,根据所有高级胶囊长度的softmax激活值判断待测试的初始语音音素数据的所属类别。

[0065] 所述初始语音音素数据是通过原始语音数据经过预处理得到,具体包括以下步骤:

[0066] A.对原始语音进行带通滤波平滑处理,去除毛刺噪声点,公式如下:

[0067] $s_p=filter_{a,b}(s_r)$

[0068] 其中,s_p表示滤波后的语音信号,s_r表示未经处理的原始语音数据,$filter_{a,b}$为滤波操作,频率的通带为aHz-bHz;

[0069] B.对滤波后的语音信号进行分帧处理;本实施例按照20ms级别的颗粒度对滤波后的语音信号进行分帧处理,即每一帧的时间长度是20ms,本实施例中使用的数据的采样频率是16000Hz,所以每一帧包含320个采样点;

[0070] C.对分帧后的每一帧语音信号进行特征提取，选取w个特征作为每一帧的w维特征向量；本实施例中通过傅里叶变换计算每一帧的基音频率、共振峰频率、能量频谱倾斜以及带宽等构成w个维度作为每一帧的特征，去除无关频率成分的影响，w为25；

[0071] D.对每一帧的w维(25维)特征向量进行归一化处理得到初始语音音素数据，即将每一帧的w维特征向量按维度除以一个归一化因子，所述归一化因子是一个w维(25维)向量l_norm。将语音信号的特征按各个维度将数值归一化到0-1之间，能够加速分类模型的收敛性能。

[0072] 所述步骤11)中所述编码方法具体为全连接网络的编码、二维卷积网络的编码或者混合型编码。

[0073] 所述步骤11)利用混合型编码得到初级胶囊的具体步骤为：

[0074] 111)初始语音音素数据input0(归一化的w维(25维)特征向量)经过一个全连接层连接到具有m_1个单元的隐藏层h_1，本实施例中m_1为1600；全连接层的权重为en_w_1，偏置为en_b_1，则隐藏层h_1的输入为a_1＝input0·en_w_1+en_b_1，其中·为矩阵乘法，经过激活函数sigmoid作用得到隐藏层h_1的输出数据z_1＝sigmoid(a_1)，a_1为一个向量，sigmoid函数映射关系作用于向量的每一个元素上，本实施例中a_1的维度为1600。en_w_1和en_b_1均为可迭代更新的参数。

[0075] 112)将隐藏层h_1的输出数据z_1重塑为一个三维张量input1，即m_2个b_1*b_1大小的特征图谱，本实施例中b_1*b_1为5*5，m_2为64，可通过第三方软件实现重塑，通过卷积核大小为k_1*k_1的二维卷积层连接到具有m_3个b_2*b_2大小的特征图谱的隐藏层h_2；本实施例中m_3为256，b_2为3；卷积层的权重为en_w_2，偏置为en_b_2，且卷积核大小k_1*k_1为3*3，隐藏层h_2的输入为a_2＝input1*en_w_2+en_b_2，其中*为矩阵二维卷积，经过激活函数sigmoid作用得到隐藏层h_2的输出z_2＝sigmoid(a_2)，本实施例中，a_2的大小为3*3*256。en_w_2和en_b_2均为可迭代更新的参数。

[0076] 113)将隐藏层h_2的输出数据z_2重塑成n_1*pri_len大小的二维张量，代表n_1个初级胶囊，张量矩阵的每一行代表一个初级胶囊pri_cap，每个初级胶囊的维度是pri_len；本实施例中二维张量的大小为288*8，即n_1个(288)初级胶囊，每个初级胶囊是元素个数为8(维度)的一个向量。

[0077] 114)将每个初级胶囊按照如下公式进行squash归一化处理。

[0078] $$\mathrm{squash}(\mathrm{pri\_cap}) = \frac{\left\|\mathrm{pri\_cap}\right\|^2}{1+\left\|\mathrm{pri\_cap}\right\|^2} \times \frac{\mathrm{pri\_cap}}{\left\|\mathrm{pri\_cap}\right\|}$$

[0079] squash归一化函数的曲线如图4所示，该函数能将长度较短的胶囊迅速压缩到接近于0的长度，将长度较长的胶囊压缩到接近于1的长度。

[0080] 所述步骤12)构造胶囊网络的动态路由结构，将初级胶囊的信息传递给高级胶囊，高级胶囊的个数即为预定的分类类别数目，所有初级胶囊根据动态路由计算出n_2个高级胶囊的向量表示，本实施例中n_2为10，动态路由结构采用路由迭代，迭代次数为预设值，本实施例中迭代次数为3次，每一次迭代过程产生更能耦合高级胶囊以及初级胶囊的耦合系数，具体步骤为：

[0081] 121)将n_1个维度为pri_len的初级胶囊复制n_2份(10份)得到一个n_1*n_2*pri_len*1

(即288*10*8*1)的张量tile_pri_cap，n_2为预设的高级胶囊数目，基于张量的矩阵乘法得到中间张量p，具体公式如下：

[0082] $p=W\cdot tile\_pri\_cap$

[0083] 其中W是一个形如$n_1*n_2*w_n*pri_len$(即288*10*16*8)的权值张量，w_n为高级胶囊维度，本实施例中为16维，·代表张量的矩阵乘法，即执行n_1*n_2(即288*10)次w_n*pri_len(即16*8)与pri_len*1(即8*1)的矩阵相乘，p是形如$n_1*n_2*w_n*1$(即288*10*16*1)的动态路由的中间张量，其中W是可迭代更新的参数；

[0084] 122)进行动态路由的迭代，根据柔性最大值softmax规则归一化张量B的每一行得到激活后的耦合系数张量C，耦合系数张量C决定了每个初级胶囊的信息传递到每个高级胶囊的比例，耦合系数张量C中的C_{ij}代表中间张量p中第i个初级胶囊的信息分配到第j个高级胶囊的比例，初次迭代时，张量B为形如n_1*n_2(即288*10)的全零张量；

[0085] 将激活后的耦合系数张量C根据张量乘法乘以中间张量p，具体公式如下：

[0086] $S=multiply(C,p)$

[0087] 其中multiply代表p中的n_1*n_2个(即288*10个)w_n维(16维)向量按对应位置乘以C中的n_1*n_2个(即288*10个)耦合系数，S是形如$n_1*n_2*w_n*1$(即288*10*16*1)的媒介张量，代表经过耦合系数进行初级胶囊和高级胶囊之间信息传递后的中间信息媒介；

[0088] 123)将媒介张量S按照第一维度进行求和操作并且保留维度，将第一个维度轴上的n_1个(288个)数字相加，得到n_2个(10个)维度是w_n(即16)的高级胶囊，再使用squash函数对高级胶囊进行归一化处理，得到形如$1*n_2*w_n$(即1*10*16)的张量，如果本次迭代为最后一次动态路由的迭代(本实施例中即第三次迭代)，则输出归一化后的高级胶囊pho_cap，进入步骤13)；如果本次迭代不是最后一次迭代，则进入步骤124)；

[0089] 124)将归一化后的高级胶囊pho_cap按照第一个维度复制n_1份(288份)，n_1为初级胶囊个数，得到形如$n_1*n_2*w_n$(即288*10*16)的张量v_j，将张量v_j扩增维度得到形如$n_1*n_2*1*w_n$(即288*10*1*16)的张量，按张量的矩阵乘法计算张量v_j与中间张量p的乘积得到形如n_1*n_2*1的张量胶囊的相似性度量矩阵(即是进行了288*10次1*16的矩阵与16*1的矩阵的相乘运算)，将相似性度量矩阵张量按元素对应加到张量B上，至此完成一次动态路由的迭代过程，进入步骤122)；

[0090] 所述步骤2)构造胶囊网络的解码器，解码器使用堆叠的三层全连接网络构成，将真实音素符号对应的高级胶囊进行解码重构具体步骤为：

[0091] 21)为了屏蔽无效信息的干扰，将非真实音素符号对应的高级胶囊中的数据置零，并将n_2个维度w_n为高级胶囊的数据重塑成n_2*w_n维度的向量v_pho_cap；本实施例中为10*16维度的向量v_pho_cap；

[0092] 22)将向量v_pho_cap经过包括隐藏层de_fc1和隐藏层de_fc2的全连接层，隐藏层de_fc1的单元数量为m_4，隐藏层de_fc2的单元数量为m_5，本实施例中m_4、m_5分别为128和64，以全连接的方式连接到输出维度为w(即25)的解码器输出层，w为初始语音音素数据的维度，解码出的重构数据与初始语音音素数据具有相同的数据格式。

[0093] 所述步骤3)基于预设的损失函数得到总损失，对胶囊网络的参数进行优化，最小化总损失，具体步骤如下：

[0094] 31)采用解码器输出的重构数据与初始语音音素数据之间距离的平方差作为重构

损失L_r;重构损失指示了解码器输出与输入的误差,具体公式如下:

[0095] s_d=decode(pho_cap)

[0096] L_r=(s_d-input0)^2

[0097] 其中s_d表示根据高级胶囊pho_cap解码重构出的数据,decode表示解码器的解码过程,input0表示初始语音音素数据。

[0098] 32)根据所有高级胶囊的长度得到分类损失L_c,分类损失表示高级胶囊的长度与期望值阈值的差距,计算公式如下:

[0099] $$L\_c=\sum_k T_k \max\left(0, m^+-\left\|pho\_cap_k\right\|\right)^2+\lambda\left(1-T_k\right)\max\left(0, \left\|pho\_cap_k\right\|-m^-\right)^2$$

[0100] 其中k是预测音素符号,T_k是一个分类标签指示函数,当k指向真实音素符号时,T_k为1,||pho_cap_k||表示预测音素符号k对应的归一化后的高级胶囊的长度,m^+和m^-分别为归一化后的单个高级胶囊长度的上下限值,max函数表示取两个数值中较大的一个,即当真实音素符号对应的高级胶囊长度超过上限,则该高级胶囊的分类损失忽略,非真实音素符号对应的高级胶囊长度低于下限时,则该高级胶囊的分类损失忽略,λ为非真实音素符号对应的分类损失之和的缩放系数;

[0101] 33)总损失函数L_t=L_c+η·L_r,其中η是用于调整分类损失和重构损失在总损失中的比重的比例系数,采用自适应矩估计法对总损失函数进行优化,即根据链式求导以及反向传播法则求出总损失函数对胶囊网络中每一个可更新参数的导数,进而最小化总损失,优化胶囊网络。

[0102] 所述步骤33)中最小化总损失具体为:采用小批量的梯度下降法,在训练时每次前向计算的总损失为小批量数据中所有输入数据对应的总损失的平均值;使用自适应矩估计法对胶囊网络中所有可更新参数进行求导,并使用梯度下降法对每个小批量训练数据更新胶囊网络中的可更新参数。本实施例中自适应矩估计法的学习率设定为0.001,本实施例的分类方法能够收敛到较优的解,训练时间长短由具体的训练数据大小决定,一般设定训练数据中每个样本都被计算50次后停止训练。

[0103] 在训练完后,参数都已固定,在测试阶段不需要再计算胶囊对于真实音素符号的损失,故将计算损失函数的部分以及解码重构部分截断去除,直接根据所有高级胶囊长度的softmax激活值来判断输入语音数据的所属类别。

[0104] 采用数据集Arctic对本发明的分类方法进行测试,该数据集采集了两名发音标准的受试者共2264个句子的音频文件,其中音频文件的采样频率是16kHz,2264个句子一共包含了40个类别的音素。为测试本发明在不同分类复杂度下的性能构建了四种测试场景:场景一中分类的音素对象是l和n,共两类;场景二中分类的音素对象是er,ey和eh,共三类;场景一中分类的音素对象是ao,ae,ax和ah,共四类;场景四中分类的音素对象是b,d,f,g,k,l,n,p,r和s,共十类。具体测试场景如表1所示,测试结果如表2所示。

[0105] 表1四种测试场景

[0106]

	采集对象人数	采样频率	句子数	音素对象
场景一	男1女1	16kHz	2264	l,n
场景二	男1女1	16kHz	2264	er,ey,eh

场景三	男1女1	16kHz	2264	ao,ae,ax,ah
场景四	男1女1	16kHz	2264	b,d,f,g,k,l,n,p,r,s

[0107] 表2四种测试场景中的测试结果

[0108]

	训练正确率	测试正确率
场景一	100%	99.46%
场景二	100%	97.77%
场景三	100%	97.86%
场景四	100%	88.22%

[0109]

[0110] 采用对音素符号的分类的正确率作为评价指标，正确率的计算分为训练阶段和测试阶段，训练正确率表体现了本发明对训练数据的所对应真实音素符号的预测准确率，测试正确率体现本发明对新数据的泛化能力。

[0111] 由上表可知，本发明提出的分类方法对所有训练数据都具有完美的拟合能力，即使是对于从未被该分类方法接触过的测试数据也具有良好的泛化能力。

[0112] 本领域内的技术人员可以对本发明进行改动或变型的设计但不脱离本发明的思想和范围。因此，如果本发明的这些修改和变型属于本发明权利要求及其等同的技术范围之内，则本发明也意图包含这些改动和变型在内。

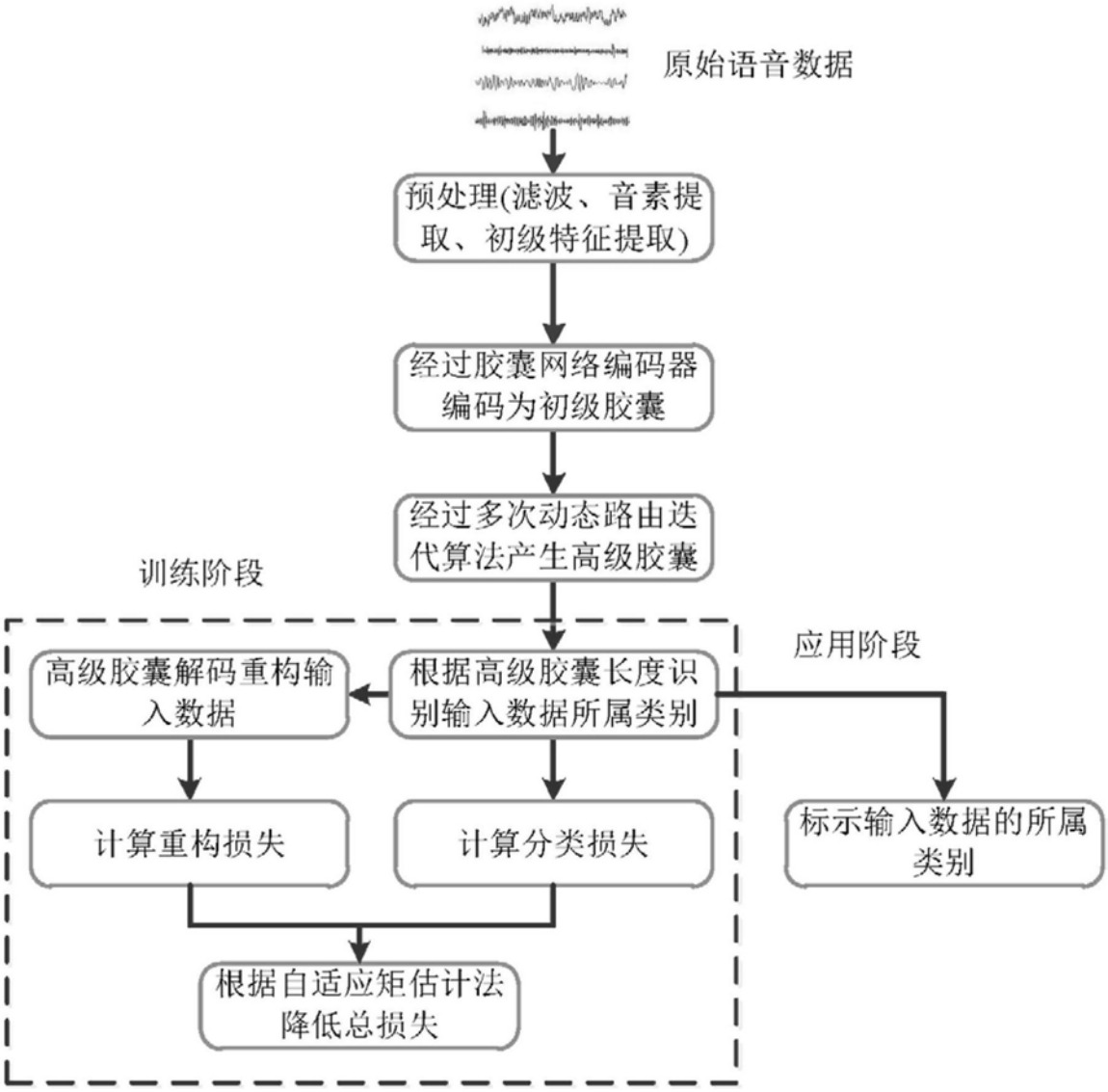

原始语音数据
预处理(滤波、音素提取、初级特征提取)
经过胶囊网络编码器编码为初级胶囊
经过多次动态路由迭代算法产生高级胶囊
训练阶段
应用阶段
高级胶囊解码重构输入数据
根据高级胶囊长度识别输入数据所属类别
计算重构损失
计算分类损失
标示输入数据的所属类别
根据自适应矩估计法降低总损失

图1

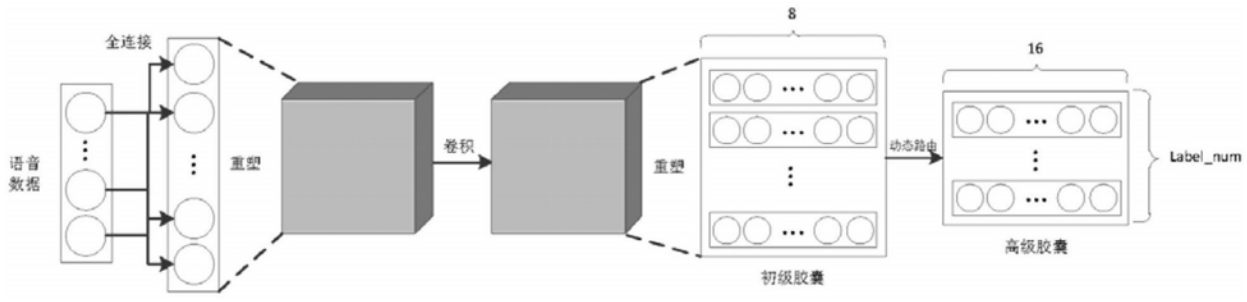

全连接
语音数据
重塑
卷积
重塑
8
初级胶囊
动态路由
16
Label_num
高级胶囊

图2

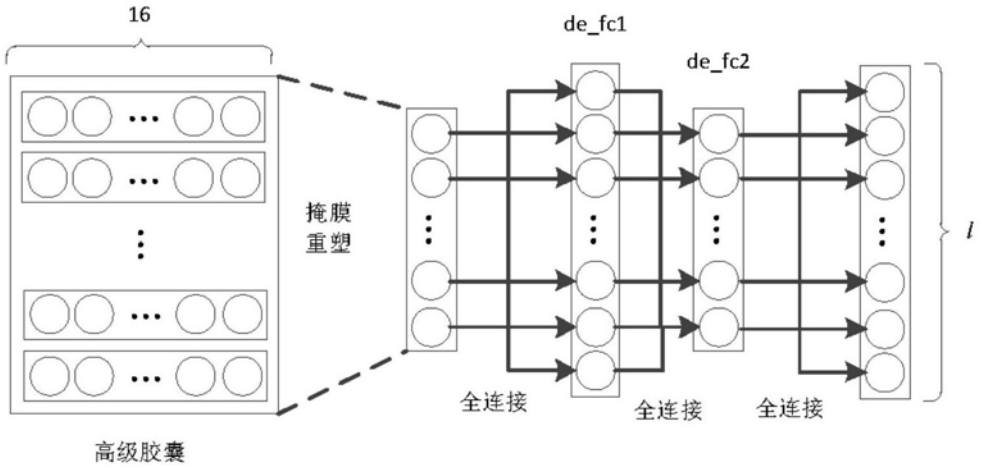

16
de_fc1
de_fc2
掩膜
重塑
全连接
全连接
全连接
l
高级胶囊

图3

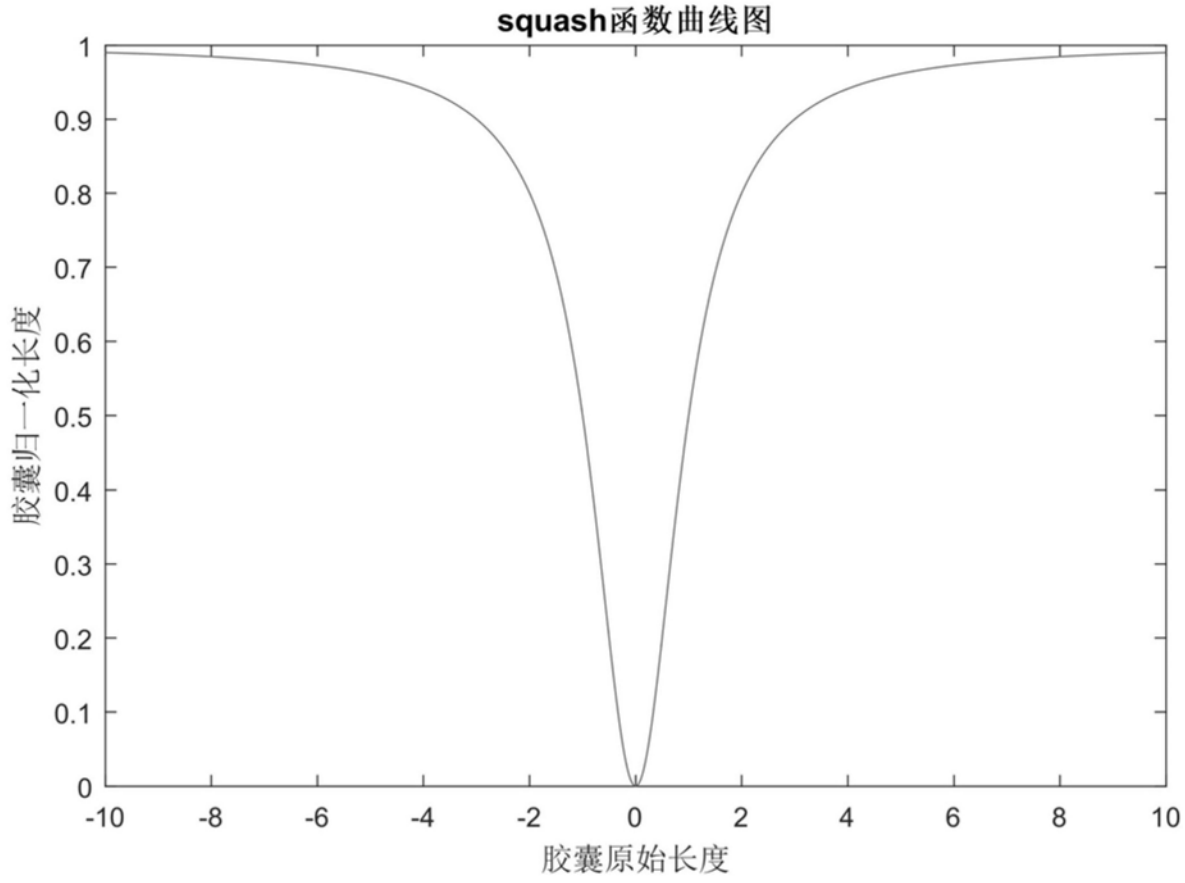

squash函数曲线图
胶囊归一化长度
胶囊原始长度
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-10
-8
-6
-4
-2
0
2
4
6
8
10

图4