(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 109979436 B
(45)授权公告日 2020.11.13

(21)申请号 201910294272.X

(22)申请日 2019.04.12

(65)同一申请的已公布的文献号
申请公布号 CN 109979436 A

(43)申请公布日 2019.07.05

(73)专利权人 南京工程学院
地址 211167 江苏省南京市江宁科学园弘景大道1号

(72)发明人 陈巍 尹伊琳

(74)专利代理机构 重庆宏知亿知识产权代理事务所(特殊普通合伙) 50260
代理人 梁山丹

(51)Int.Cl.
G10L 15/02(2006.01)
G10L 15/06(2013.01)
G10L 15/08(2006.01)
G10L 15/16(2006.01)
G10L 15/26(2006.01)
G10L 19/02(2013.01)
G10L 25/63(2013.01)

(56)对比文件
CN 103514879 A,2014.01.15
CN 106683666 A,2017.05.17
CN 203552694 U,2014.04.16
CN 109065034 A,2018.12.21
CN 101858938 A,2010.10.13
CN 104538027 A,2015.04.22
CN 108701452 A,2018.10.23
CN 102800316 A,2012.11.28
US 2018061397 A1,2018.03.01
张稳.基于神经网络的语音识别系统的实现.《中国优秀硕士学位论文全文数据库》.2013,(第12期),第5-55页.
Jan Zwlinka etc.Neural-Network-Based Spectrum Processing for Speech Recognition and Speaker Verification.《International Conference on Statistical Language and Speech Processing》.2015,第288-299页.

审查员 李春雨

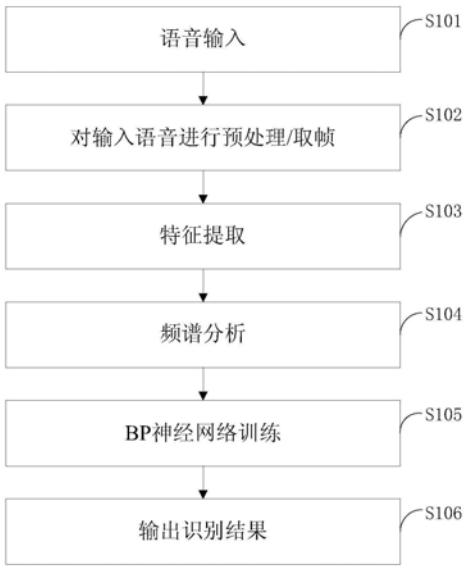
权利要求书2页 说明书9页 附图3页

(54)发明名称

一种基于频谱自适应法的BP神经网络语音识别系统及方法

(57)摘要

本发明属于语音识别技术领域,公开了一种基于频谱自适应法的BP神经网络语音识别系统及方法,基于频谱自适应法的BP神经网络语音识别方法包括:语音输入、对输入语音进行预处理/取帧、特征提取、频谱分析、BP神经网络训练、输出识别结果。本发明利用声学特征表征语音内容,不依赖于说话者或词汇内容,将韵律和音质特征整合到系统中;引入频谱变换自适应法补偿三种失真源(扬声器的差异,录音通道的变化和嘈杂环境)、重建训练向量和测试向量之间的正确相关性;通过BP神经网络算法对机器进行静态训练,进而令识别参数不断逼近最佳状态,提高识别率。



CN 109979436 B

1.一种基于频谱自适应法的BP神经网络语音识别方法,其特征在于,所述基于频谱自适应法的BP神经网络语音识别方法包括:

步骤一,语音输入;

步骤二,对输入语音进行预处理/取帧;

步骤三,特征提取;

步骤四,频谱分析;

步骤五,BP神经网络训练;

步骤六,输出识别结果;

步骤四频谱分析采用频谱自适应算法;频谱自适应算法包括:

令训练向量和测试向量分别是向量X(1)和X(2),假设:

$U=AX^{(1)},V=BX^{(2)}$ (1)

其中A和B是对应于X(1)和X(2)的变换矩阵,u和v是参考空间中公式(1)x和(2)x的映射;将均方误差最小化:

$D=E\{(U-V)^2\}$,其中$U=AX^{(1)},V=BX^{(2)}$ (2)

带约束$E\{U^2\}=E\{V^2\}=1$;做U和V的最大相关,u和v在当时不为零;

$X=\begin{bmatrix}X^{(1)}\\X^{(2)}\end{bmatrix}$;假设语音倒谱的长期均值为零,令$E\{X\}=0$,分别从训练向量和测试向量中减去信道特征;得到的$E\{X^{(1)}\}=E\{X^{(2)}\}=0$,$X^{(1)}=X^{(1)}-\overline{X}^{(1)}$和$X^{(2)}=X^{(2)}-\overline{X}^{(2)}$,得到相关矩阵:

$$\sum=\begin{bmatrix}\sum_{11} & \sum_{12}\\ \sum_{21} & \sum_{22}\end{bmatrix} \quad (3)$$

得到关系:

$I=E\{U^2\}=E\{A'X^{(1)}X^{(1)'}A\}=A'\Sigma_{11}A$ (4)

$I=E\{V^2\}=E\{B'X^{(2)}X^{(2)'}B\}=B'\Sigma_{22}B$ (5)

$E\{U\}=E\{A'X^{(1)}\}=A'E\{X^{(1)}\}=0$ (6)

$E\{V\}=E\{B'X^{(2)}\}=B'E\{X^{(2)}\}=0$ (7)

$E\{UV\}=E\{A'X^{(1)}X^{(2)}B\}=A'\Sigma_{12}B$ (8)

问题改写为:

$$\psi=A'\sum_{12}B-\lambda(A'\sum_{11}A-1)-\frac{1}{2}\mu(B'\sum_{22}B-1) \quad (9)$$

令$\frac{\partial\psi}{\partial A}=0,\frac{\partial\psi}{\partial B}=0$,得到

$$\begin{bmatrix}-\lambda\sum_{11} & \sum_{12}\\ \sum_{21} & -\lambda\sum_{22}\end{bmatrix}\begin{bmatrix}A\\B\end{bmatrix}=0 \quad (10)$$

满足

$$\begin{bmatrix} -\lambda\sum_{11} & \sum_{12} \\ \sum_{21} & -\lambda\sum_{22} \end{bmatrix} = 0 \qquad (11)$$

特征向量$(a^{(1)},b^{(1)}),(a^{(2)},b^{(2)}),.....(a^{(p)},b^{(p)})$对应于$\lambda_1,\lambda_2,...\lambda_p$是转换矩阵A和B的行向量;通过计算将测试向量$B^{-1}A(X^{(2)}-\overline{X}^{(2)})+\overline{X}^{(1)}$映射到训练空间。

2.如权利要求1所述基于频谱自适应法的BP神经网络语音识别方法,其特征在于,所述步骤三特征提取具体包括:

语音输入即把语音输入设备采集的语音进行原始输入,通过扩音器将未知声音转化为电信号输入识别系统,进行预处理;预处理包括采样语音信号、反混叠带通滤波、去除个体发音差异和设备、环境引起的噪声影响,并且每隔一定时间间隔取出部分信号处理,确定帧的尺寸以及计算重叠率;根据取帧划分的语音信号的每帧中提取出韵律特征和质量特征,确定特征集中最佳分类的特征;在BP神经网络训练阶段,对特征进行分析并得到信号归属词汇,为每个词条建立一个模型,保存为模板库;在识别阶段,使用所获得的特征集来执行情感识别,语音信号经过相同的通道得到语音特征参数,生成测试模板,与参考模板进行匹配,生成识别结果。

3.如权利要求1所述基于频谱自适应法的BP神经网络语音识别方法,其特征在于,所述步骤五BP神经网络训练包括:采用输入层、隐藏层、输出层三层结构作为情感识别的框架;

输入神经元的数量=特征数量;

隐藏层数量=(特征数量+情感数量)/2;

输出神经元数量=情感数量。

4.一种实施权利要求1所述基于频谱自适应法的BP神经网络语音识别方法的基于频谱自适应法的BP神经网络语音识别控制系统。

一种基于频谱自适应法的BP神经网络语音识别系统及方法

技术领域

[0001] 本发明属于语音识别技术领域，尤其涉及一种基于频谱自适应法的BP神经网络语音识别系统及方法。

背景技术

[0002] 目前，最接近的现有技术：

[0003] 特征参数匹配法、隐马尔可夫法和神经网络法。现有语音识别技术多有环境噪声影响、说话人距离和位置变化的影响以及说话人心理和生理变化的影响等，缺乏稳定性和自适应性。

[0004] 语音识别的应用往往工作环境复杂，声学特征的精确提取通常较难获得。这就需要语音识别系统具有一定的自适应性并进行BP算法训练。目前，常被用于语音识别技术的方法有HMM模型，BP神经网络算法。

[0005] 然而，当周围存在较多高频噪声或说话人因情感变化而使说话口吻改变时，系统识别性能减弱，造成语音识别率不够。随科技发展，计算机和机器人需具有更强的表达、识别和理解能力，从而人机界面更为高效。

[0006] 综上所述，现有技术存在的问题是：现有语音识别技术多有环境噪声影响、说话人距离和位置变化的影响以及说话人心理和生理变化的影响等，缺乏稳定性和自适应性。

[0007] 解决上述技术问题的难度：任务过程中因环境变化、说话人距离改变、说话人因情感变化而改变说话口吻从而影响所提取特征值的有效性；任务过程中因扬声器的差异、录音通道的变化从而产生训练条件与测试条件间的不匹配；任务过程中因建立数据库差异导致某些语言无法识别等。

[0008] 解决上述技术问题的意义：基于频谱自适应法的BP神经网络语音识别方法，用以提高训练条件与测试条件间的匹配程度；利用BP神经网络算法对机器训练，进而令识别参数不断逼近最佳状态，提高识别率。

发明内容

[0009] 针对现有技术存在的问题，本发明提供了一种基于频谱自适应法的BP神经网络语音识别方法。

[0010] 本发明是这样实现的，一种基于频谱自适应法的BP神经网络语音识别方法，包括：

[0011] 步骤一，语音输入；

[0012] 步骤二，对输入语音进行预处理/取帧；

[0013] 步骤三，特征提取；

[0014] 步骤四，频谱分析；

[0015] 步骤五，BP神经网络训练；

[0016] 步骤六，输出识别结果。

[0017] 进一步，所述步骤三特征提取具体包括：

[0018] 语音输入即把语音输入设备采集的语音进行原始输入，通过扩音器将未知声音转化为电信号输入识别系统，进行预处理；预处理包括采样语音信号、反混叠带通滤波、去除个体发音差异和设备、环境引起的噪声影响等，并且每隔一定时间间隔取出部分信号处理，确定帧的尺寸以及计算重叠率；根据取帧划分的语音信号的每帧中提取出韵律特征和质量特征，确定特征集中最佳分类的特征；在BP神经网络训练阶段，主要是对特征进行分析并得到信号归属词汇，为每个词条建立一个模型，保存为模板库。在识别阶段，使用所获得的特征集来执行情感识别，语音信号经过相同的通道得到语音特征参数，生成测试模板，与参考模板进行匹配，基于本专利算法规则生成识别结果。

[0019] 进一步，步骤四频谱分析采用频谱自适应算法；频谱自适应算法包括：

[0020] 令训练向量和测试向量分别是向量X(1)和X(2)，假设：

[0021] $U=AX^{(1)},V=BX^{(2)}$ (1)

[0022] 其中A和B是对应于X(1)和X(2)的变换矩阵，u和v是参考空间中公式(1)x和(2)x的映射；将均方误差最小化：

[0023] $D=E\{(U-V)^2\}$，其中$U=AX^{(1)},V=BX^{(2)}$ (2)

[0024] 带约束$E\{U^2\}=E\{V^2\}=1$；做U和V的最大相关，u和v在当时不为零；

[0025] $X=\begin{bmatrix}X^{(1)}\\X^{(2)}\end{bmatrix}$；假设语音倒谱的长期均值为零，令$E\{X\}=0$，分别从训练向量和测试向量中减去信道特征；得到的$E\{X^{(1)}\}=E\{X^{(2)}\}=0$，$X^{(1)}=X^{(1)}-\overline{X}^{(1)}$和$X^{(2)}=X^{(2)}-\overline{X}^{(2)}$，得到相关矩阵：

[0026] $$\sum=\begin{bmatrix}\sum_{11} & \sum_{12}\\ \sum_{21} & \sum_{22}\end{bmatrix} \quad (3)$$

[0027] 得到关系：

[0028] $I=E\{U^2\}=E\{A'X^{(1)}X^{(1)}'A\}=A'\sum_{11}A$ (4)

[0029] $I=E\{V^2\}=E\{B'X^{(2)}X^{(2)}'B\}=B'\sum_{22}B$ (5)

[0030] $E\{U\}=E\{A'X^{(1)}\}=A'E\{X^{(1)}\}=0$ (6)

[0031] $E\{V\}=E\{B'X^{(2)}\}=B'E\{X^{(2)}\}=0$ (7)

[0032] $E\{UV\}=E\{A'X^{(1)}X^{(2)}B\}=A'\sum_{12}B$ (8)

[0033] 问题改写为：

[0034] $$\psi=A'\sum_{12}B-\lambda(A'\sum_{11}A-1)-\frac{1}{2}\mu(B'\sum_{22}B-1) \quad (9)$$

[0035] 令$\frac{\partial\psi}{\partial A}=0,\frac{\partial\psi}{\partial B}=0$，得到

[0036] $$\begin{bmatrix}-\lambda\sum_{11} & \sum_{12}\\ \sum_{21} & -\lambda\sum_{22}\end{bmatrix}\begin{bmatrix}A\\B\end{bmatrix}=0 \quad (10)$$

[0037] 满足

[0038] $$\begin{bmatrix} -\lambda\sum_{11} & \sum_{12} \\ \sum_{21} & -\lambda\sum_{22} \end{bmatrix} = 0 \qquad (11)$$

[0039] 特征向量$(a^{(1)},b^{(1)})$,$(a^{(2)},b^{(2)})$,……$(a^{(p)},b^{(p)})$对应于$\lambda_1,\lambda_2,\cdots\lambda_p$是转换矩阵A和B的行向量;通过计算将测试向量$B^{-1}A(X^{(2)}-\overline{X}^{(2)})+\overline{X}^{(1)}$映射到训练空间。

[0040] 进一步,所述步骤五BP神经网络训练包括:采用输入层、隐藏层、输出层三层结构作为情感识别的框架;

[0041] 输入神经元的数量=特征数量;

[0042] 隐藏层数量=(特征数量+情感数量)/2;

[0043] 输出神经元数量=情感数量。

[0044] BP神经网络训练包括:反向传播神经网络(BPNN)即BP网络,BPNN原则上以多层感知(MLP)为系统框架,以反向传播算法为训练规则。MLP即多层感知器,是一种前向结构的人工神经网络,通常使用静态反向传播进行训练,对静态模式进行分类。该网络可以手动构建,在训练期间也可以监视和修改网络。MLP模型中的多层结构表明它由多层神经元组成。另外,两层神经元之间的信号传递模式与单层神经元相同。

[0045] 本发明的另一目的在于提供一种基于频谱自适应法的BP神经网络语音识别控制系统。

[0046] 综上所述,本发明的优点及积极效果为:

[0047] 本发明成功对七种离散的情感状态(愤怒、厌恶、恐惧、快乐、中立、悲伤、惊讶)识别。在10dB信噪比下,以16kHz的采样率,用中文记录了7位发言者的情感语音数据库,每种情感用100个语音进行训练。

[0048] 而一组每种情感100个话语的分离被用来测试。

[0049] 对比实验结果如图5所示,“1”代表愤怒,“2”代表厌恶,“3”代表恐惧,“4”代表欢乐,“5”代表中立,“6”代表悲伤,“7”代表惊奇。

[0050] 频谱自适应法和BP神经网络法不仅提高了识别率,而且在低信噪比情况下也提高了系统的鲁棒性,这说明频谱自适应法很好地补偿了训练集和测试集之间的不匹配,用频谱自适应法作为补偿比不用频谱自适应法更好。如图5所示。其次,本发明使用了男性语言数据库。利用DB8小波对神经网络进行了13级分解后的特征向量训练,对神经网络进行了识别四种不同情感的测试,模糊矩阵中的识别精度如表1所示。本发明可获得72.055%的整体识别精度,解决了语音识别技术的情感识别这一难题。

[0051] 表1

[0052]

情感分类	中性	快乐	悲伤	生气
中性	76.47%	17.64%	5.88%	0%
快乐	17.64%	52.94%	17.6%	11.76%
悲伤	17.64%	11.76%	70.58%	0%
生气	11.76%	0%	0%	88.23%

。

[0053] 本发明述及方法利用声学特征,该特征有效表征语音内容,不依赖于说话者或词汇内容,并将韵律和音质特征整合到系统中;采用离散小波变换进行性别分析;利用统一的

频谱变换自适应法补偿三种失真源(扬声器的差异,录音通道的变化和嘈杂环境)、重建训练向量和测试向量之间的正确相关性;通过BP神经网络算法对机器进行静态训练,进而令识别参数不断逼近最佳状态,提高识别率。

附图说明

[0054] 图1是本发明实施例提供的基于频谱自适应法的BP神经网络语音识别方法流程图。

[0055] 图2是本发明实施例提供的语音识别过程图。

[0056] 图3是本发明实施例提供的频谱自适应算法计算流程图。

[0057] 图4是本发明实施例提供的三层神经网络框架图。

[0058] 图5是本发明实施例提供的不同情感的识别错误率图。

具体实施方式

[0059] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合实施例,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仅用以解释本发明,并不用于限定本发明。

[0060] 现有技术的语音识别中,没有用以提高训练条件与测试条件间的匹配程度;利用BP神经网络算法对机器训练,进而令识别参数不断逼近最佳状态,造成识别率低。

[0061] 解决上述问题,下面结合具体方案对本发明作详细描述。

[0062] 本发明采用神经网络作为语音识别平台,提出一种提高语音识别率的识别方法,采用频谱自适应算法以提高训练条件与测试条件间的匹配程度;利用BP神经网络算法对机器训练,令识别参数不断逼近最佳状态。

[0063] 如图1所示,本发明实施例提供的基于频谱自适应法的BP神经网络语音识别方法包括:

[0064] S101:语音输入。

[0065] S102:对输入语音进行预处理/取帧。

[0066] S103:特征提取。

[0067] S104:频谱分析。

[0068] S105:BP神经网络训练。

[0069] S106:输出识别结果。

[0070] 所述步骤S103的特征提取具体包括:

[0071] 特征提取既是大幅压缩信息量的过程,也是信号解卷的过程。将语音信号转换成一组特征矢量序列,使模式划分器能更好地划分。由于语音信号是非平稳信号,本发明假设在非常短的时间间隔内信号静止,即在此时间间隔内信号稳定,因此可每隔一定间隔取出部分信号处理。确定帧的尺寸以及计算重叠率称为取帧,计算重叠率即强化从一帧到另一帧的转换以防止信息丢失。在该阶段,根据取帧划分的语音信号的每帧中提取出韵律特征和质量特征。特征集中的单位差异和数据的数字大小直接影响分类器的性能,采用标准化技术克服该影响;特征选择法用于确定将从特征集中,实现最佳分类的特征。最后,使用所获得的特征集来执行情感识别。

[0072] 所述步骤S104的频谱分析采用频谱自适应算法。所述频谱自适应算法包括：

[0073] 频谱自适应算法是一种指数平滑预测方法，可用于非平稳时间序列的预测。预处理语音信号可以表示为一系列特征向量，每个向量可以被认为是特征向量空间中的一个点，从而运用频谱自适应算法，改善训练向量和测试向量之间的差异并进行补偿，该方法没有直接将测试空间转换为训练空间，它使得训练向量和测试向量在参考空间(第三空间)中的相关性最大。令训练向量和测试向量分别是向量X(1)和X(2)，可以假设：

[0074] $U=AX^{(1)}, V=BX^{(2)}$ (1)

[0075] 其中A和B是对应于X(1)和X(2)的变换矩阵，u和v是参考空间中(1)x和(2)x的映射。将均方误差最小化：

[0076] $D=E\{(U-V)^2\}$ (2)

[0077] 带约束$E\{U^2\}=E\{V^2\}=1$。做U和V的最大相关，并保证u和v在当时不能为零。通过以下步骤：如图3所示。

[0078] 假设$X=\begin{bmatrix}X^{(1)}\\X^{(2)}\end{bmatrix}$。假设语音倒谱的长期均值为零，可以令$E\{X\}=0$，分别从训练向量和测试向量中减去信道特征。可以得到的$E\{X^{(1)}\}=E\{X^{(2)}\}=0$，$X^{(1)}=X^{(1)}-\overline{X}^{(1)}$和$X^{(2)}=X^{(2)}-\overline{X}^{(2)}$，因此得到相关矩阵：

[0079] $$\sum=\begin{bmatrix}\sum_{11} & \sum_{12}\\ \sum_{21} & \sum_{22}\end{bmatrix} \quad (3)$$

[0080] 得到关系：

[0081] $I=E\{U^2\}=E\{A'X^{(1)}X^{(1)}'A\}=A'\Sigma_{11}A$ (4)

[0082] $I=E\{V^2\}=E\{B'X^{(2)}X^{(2)}'B\}=B'\Sigma_{22}B$ (5)

[0083] $E\{U\}=E\{A'X^{(1)}\}=A'E\{X^{(1)}\}=0$ (6)

[0084] $E\{V\}=E\{B'X^{(2)}\}=B'E\{X^{(2)}\}=0$ (7)

[0085] $E\{UV\}=E\{A'X^{(1)}X^{(2)}B\}=A'\Sigma_{12}B$ (8)

[0086] 问题可以改写为：

[0087] $$\psi=A'\sum_{12}B-\lambda(A'\sum_{11}A-1)-\frac{1}{2}\mu(B'\sum_{22}B-1) \quad (9)$$

[0088] 如果令$\frac{\partial\psi}{\partial A}=0,\frac{\partial\psi}{\partial B}=0$，得到

[0089] $$\begin{bmatrix}-\lambda\sum_{11} & \sum_{12}\\ \sum_{21} & -\lambda\sum_{22}\end{bmatrix}\begin{bmatrix}A\\B\end{bmatrix}=0 \quad (10)$$

[0090] 必须满足

[0091] $$\begin{bmatrix}-\lambda\sum_{11} & \sum_{12}\\ \sum_{21} & -\lambda\sum_{22}\end{bmatrix}=0 \quad (11)$$

[0092] 证明方程(11)有根$\lambda_1,\lambda_2,\cdots,\lambda_P$，求解方程(11)，将典型相关问题转化为一般特征

值问题。特征向量$(a^{(1)},b^{(1)})$,$(a^{(2)},b^{(2)})$,……$(a^{(p)},b^{(p)})$对应于$\lambda_1,\lambda_2,\cdots\lambda_p$是转换矩阵A和B的行向量。最后通过计算将测试向量$B^{-1}A(X^{(2)}-\overline{X}^{(2)})+\overline{X}^{(1)}$映射到训练空间。

[0093] 所述步骤S105的BP神经网络训练包括:反向传播神经网络(BPNN)即BP网络,BPNN原则上以多层感知(MLP)为系统框架,以反向传播算法为训练规则。MLP即多层感知器,是一种前向结构的人工神经网络,通常使用静态反向传播进行训练,对静态模式进行分类。该网络可以手动构建,在训练期间也可以监视和修改网络。MLP模型中的多层结构表明它由多层神经元组成。另外,两层神经元之间的信号传递模式与单层神经元相同。本发明采用三层结构(输入层、隐藏层、输出层)作为情感识别的框架。框架如图4所示。该模型中:

[0094] 输入神经元的数量=特征数量;

[0095] 隐藏层数量=(特征数量+情感数量)/2;

[0096] 输出神经元数量=情感数量。

[0097] 本发明将韵律和音质特征整合到系统中,利用频谱自适应算法补偿三种失真源、重建训练向量和测试向量之间的正确相关性;通过BP神经网络算法对机器进行静态训练,令识别参数不断逼近最佳状态。

[0098] 本发明基于频谱自适应算法的BP神经网络语音识别方法,可以提高训练条件与测试条件间的匹配程度;利用BP神经网络算法对机器训练,进而令识别参数不断逼近最佳状态,提高识别率。

[0099] 下面结合具体实施例对本发明作进一步描述。

[0100] 实施例:

[0101] 本发明实施例提供的基于频谱自适应法的BP神经网络语音识别方法包括以下步骤:

[0102] (1)特征提取

[0103] 特征提取既是大幅压缩信息量的过程,也是信号解卷的过程。将语音信号转换成一组特征矢量序列,使模式划分器能更好地划分。由于语音信号是非平稳信号,本发明假设在非常短的时间间隔内信号静止,即在此时间间隔内信号稳定,因此可每隔一定间隔取出部分信号处理。确定帧的尺寸以及计算重叠率称为取帧,计算重叠率即强化从一帧到另一帧的转换以防止信息丢失。(帧的大小在20ms到40ms之间,重叠率为50%)在该阶段,根据取帧划分的语音信号的每帧中提取出韵律特征和质量特征。特征集中的单位差异和数据的数字大小直接影响分类器的性能,采用标准化技术克服该影响;特征选择法用于确定将从特征集中实现最佳分类的特征。通过选择特征,减小特征数据集的大小以试图提高分类性能和准确性。最后,使用所获得的特征集来执行情感识别。

[0104] 1)韵律特点

[0105] 使用一组37个特征,其中26个特征是对数f、能量和持续时间方面的模型。对数F:最大、最小、最大和最小位置、平均值、标准差、回归系数、回归系数的均方误差,以及第一帧和最后一帧的F。

[0106] 能量:最大、最小、最大和最小位置、平均值、回归系数和回归系数的均方误差。

[0107] 持续时间方面:发声和未发声区域的数量,发声和未发声帧的数量,最长发声和未发声区域,发声和未发声帧的数量比,发声和未发声区域的数量比,发声和总帧的数量比,发声和总区域的数量比。

[0108] 2)质量特点

[0109] 情感识别方法还包括与发音精度或声道特性有关的信息,如共振峰结构。在情感表达方面,有知觉的证据表明,发声质量参数的额外重要性,即声门刺激变化产生的听觉质量。

[0110] 本发明选择了16个质量特征,描述了前三个共振峰、它们的带宽、谐波噪声比、光谱能量分布、语音与清音能量比和声门流。所有描述的质量特征都是使用语音分析软件praat获得的。

[0111] (2)频谱自适应算法

[0112] 频谱自适应算法是一种指数平滑预测方法,可用于非平稳时间序列的预测。预处理语音信号可以表示为一系列特征向量,每个向量可以被认为是特征向量空间中的一个点,从而运用频谱自适应算法,改善训练向量和测试向量之间的差异并进行补偿,该方法没有直接将测试空间转换为训练空间,它使得训练向量和测试向量在参考空间(第三空间)中的相关性最大。令训练向量和测试向量分别是向量X(1)和X(2),可以假设:

[0113] $U=AX^{(1)},V=BX^{(2)}$ (1)

[0114] 其中A和B是对应于X(1)和X(2)的变换矩阵,u和v是参考空间中(1)x和(2)x的映射。将均方误差最小化:

[0115] $D=E\{(U-V)^2\}$ (2)

[0116] 带约束$E\{U^2\}=E\{V^2\}=1$。做U和V的最大相关,并保证u和v在当时不能为零。通过以下步骤:如图3所示。

[0117] 假设$X=\begin{bmatrix}X^{(1)}\\X^{(2)}\end{bmatrix}$。假设语音倒谱的长期均值为零,可以令$E\{X\}=0$,分别从训练向量和测试向量中减去信道特征。可以得到的$E\{X^{(1)}\}=E\{X^{(2)}\}=0$,$X^{(1)}=X^{(1)}-\overline{X}^{(1)}$和$X^{(2)}=X^{(2)}-\overline{X}^{(2)}$,因此得到相关矩阵:

[0118] $$\sum=\begin{bmatrix}\sum_{11} & \sum_{12}\\ \sum_{21} & \sum_{22}\end{bmatrix} \quad (3)$$

[0119] 得到关系:

[0120] $I=E\{U^2\}=E\{A'X^{(1)}X^{(1)}'A\}=A'\Sigma_{11}A$ (4)

[0121] $I=E\{V^2\}=E\{B'X^{(2)}X^{(2)}'B\}=B'\Sigma_{22}B$ (5)

[0122] $E\{U\}=E\{A'X^{(1)}\}=A'E\{X^{(1)}\}=0$ (6)

[0123] $E\{V\}=E\{B'X^{(2)}\}=B'E\{X^{(2)}\}=0$ (7)

[0124] $E\{UV\}=E\{A'X^{(1)}X^{(2)}B\}=A'\Sigma_{12}B$ (8)

[0125] 问题可以改写为:

[0126] $$\psi=A'\sum_{12}B-\lambda(A'\sum_{11}A-1)-\frac{1}{2}\mu(B'\sum_{22}B-1) \quad (9)$$

[0127] 如果令$\frac{\partial\psi}{\partial A}=0,\frac{\partial\psi}{\partial B}=0$,得到

[0128] $$\begin{bmatrix} -\lambda\sum_{11} & \sum_{12} \\ \sum_{21} & -\lambda\sum_{22} \end{bmatrix}\begin{bmatrix} A \\ B \end{bmatrix}=0 \qquad (10)$$

[0129] 必须满足

[0130] $$\begin{bmatrix} -\lambda\sum_{11} & \sum_{12} \\ \sum_{21} & -\lambda\sum_{22} \end{bmatrix}=0 \qquad (11)$$

[0131] 本发明可以证明方程(11)有根$\lambda_1,\lambda_2,\cdots,\lambda_P$,要求解方程(11),将典型相关问题转化为一般特征值问题。特征向量$(a^{(1)},b^{(1)}),(a^{(2)},b^{(2)}),\cdots\cdots(a^{(p)},b^{(p)})$对应于$\lambda_1,\lambda_2,\cdots\lambda_p$是转换矩阵A和B的行向量。最后通过计算将测试向量$B^{-1}A(X^{(2)}-\overline{X}^{(2)})+\overline{X}^{(1)}$映射到训练空间。

[0132] 本发明经测试,发现语言转换再训练具有最佳的补偿效果。但在考虑该技术的在线应用时,没有对模型进行再训练,只将测试倒谱向量转化为训练空间进行识别。

[0133] (3)、BP神经网络训练

[0134] BPNN原则上以多层感知(MLP)为系统框架,以反向传播算法为训练规则。MLP即多层感知器,是一种前向结构的人工神经网络,通常使用静态反向传播进行训练,对静态模式进行分类。该网络可以手动构建,在训练期间也可以监视和修改网络。MLP模型中的多层结构表明它由多层神经元组成。另外,两层神经元之间的信号传递模式与单层神经元相同。

[0135] 本发明采用三层结构(输入层、隐藏层、输出层)作为情感识别的框架。框架如图4所示。该模型中:

[0136] 输入神经元的数量=特征数量;

[0137] 隐藏层数量=(特征数量+情感数量)/2;

[0138] 输出神经元数量=情感数量。

[0139] 在人工神经网络的结构中,有两种输出模式。其中一个使用二进制编码来表示输出,例如,系统有32个对应的输出到5个输出神经元。因此,输出神经元的数量减少了。另一个是一对一的输出。例如,22帧需要22个输出神经元,虽然二进制编码可以使神经元的数目最小化,但它不仅识别率低,而且与一对一模式相比,实验后难以收敛。因此,这里采用了一对一的输出。参数总共包含53个特性,因此输入层中有53个单元,输出层中有7个单元。隐层神经元的数目不能太多,否则不能收敛;如果数目太小,识别误差就大。隐层中的神经元数量用以下方程式表示:

[0140] N_no=(In_number×Out_number)1/2

[0141] 其中N_no表示隐藏层单元的数量,In_number和Out_number分别表示输入和输出层单元的数量。

[0142] 本发明实施例提供一种基于频谱自适应法的BP神经网络语音识别控制系统。

[0143] 下面结合具体实验对本发明作进一步描述。

[0144] 本发明通过实验对识别系统进行了评价。在实验中,七种离散的情感状态(愤怒、厌恶、恐惧、快乐、中立、悲伤、惊讶)在整个工作中被分类。在10dB信噪比下,以16kHz的采样率,用中文记录了7位发言者的情感语音数据库,每种情感用100个语音进行训练。

[0145] 而一组每种情感100个话语的分离被用来测试。

[0146] 对比实验结果如图5所示，“1”代表愤怒，“2”代表厌恶，“3”代表恐惧，“4”代表欢乐，“5”代表中立，“6”代表悲伤，“7”代表惊奇。

[0147] 其次，本发明使用了男性语言数据库。利用DB8小波对神经网络进行了13级分解后的特征向量训练，对神经网络进行了识别四种不同情感的测试，模糊矩阵中的识别精度如表1所示。在测试网络识别四种不同情感的同时，机器获得了最大的识别准确度，在情感愤怒的情况下，最小的识别准确度是幸福。当机器试图从四个不同的情感类别中识别出中性语言时，机器获得了76.47％的识别准确率，而机器面临17.64％的困惑，情感快乐，5.88％的困惑是悲伤，机器不再面临情感愤怒的困惑。对于快乐的情感识别，机器能达到52.94％的识别准确率，17.64％的识别准确率为中性情感，17.6％的识别准确率为悲伤情感，11.76％的识别准确率为愤怒情感。在识别情感悲伤时，机器获得70.58％的识别准确率，17.64％的识别率与情感中性相混淆，11.76％的识别率与情感悲伤相混淆，不再与情感愤怒相混淆。对于情感愤怒的识别，机器识别准确率达到88.23％，与情感中性的识别混淆率达到11.76％，在情感喜怒哀乐的情况下不再出现混淆。通过本实验，本发明可获得72.055％的整体识别精度。

[0148] 表1

[0149]

情感分类	中性	快乐	悲伤	生气
中性	76.47％	17.64％	5.88％	0％
快乐	17.64％	52.94％	17.6％	11.76％
悲伤	17.64％	11.76％	70.58％	0％
生气	11.76％	0％	0％	88.23％

。

[0150] 以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等，均应包含在本发明的保护范围之内。

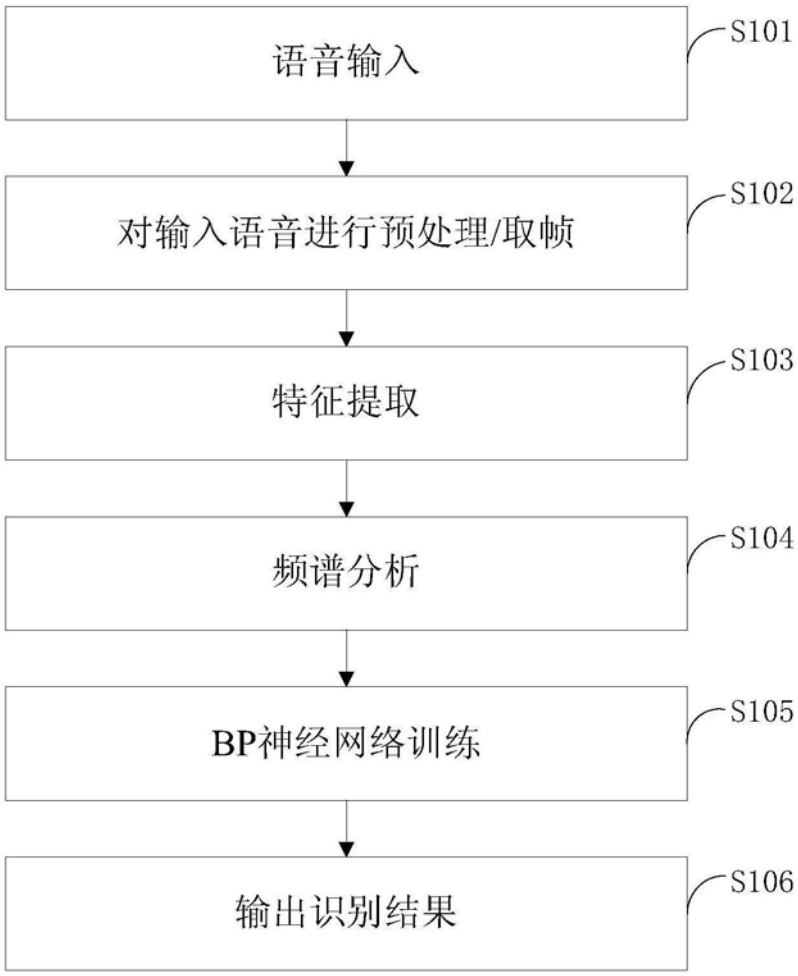

语音输入
S101
对输入语音进行预处理/取帧
S102
特征提取
S103
频谱分析
S104
BP神经网络训练
S105
输出识别结果
S106

图1

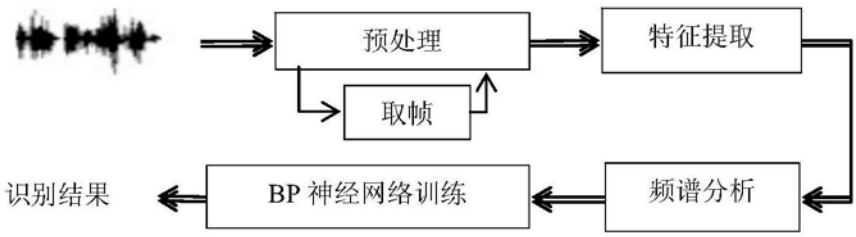

预处理
特征提取
取帧
识别结果
BP 神经网络训练
频谱分析

图2

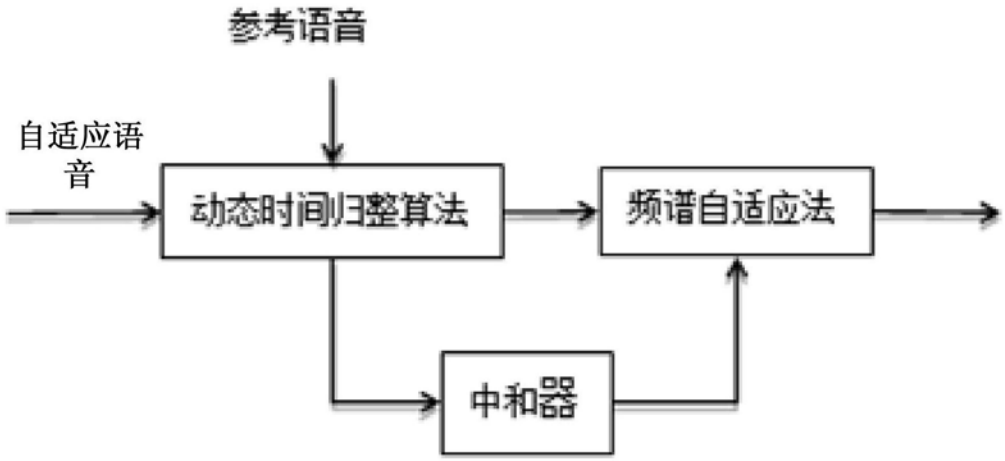


图3

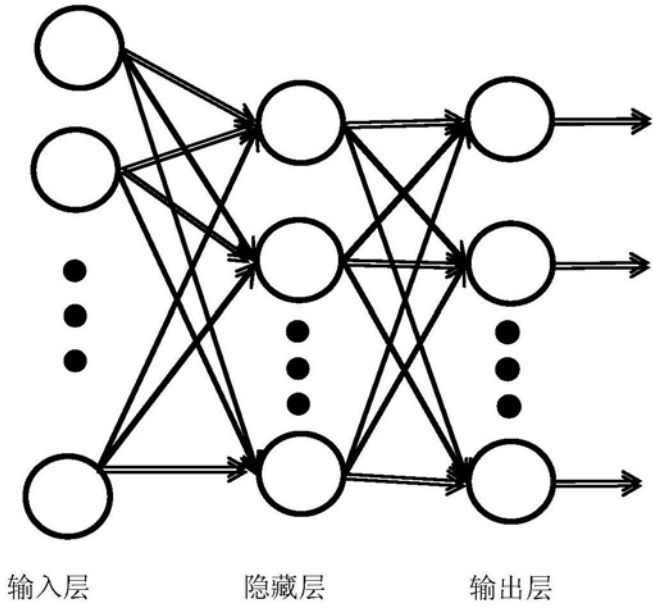


图4

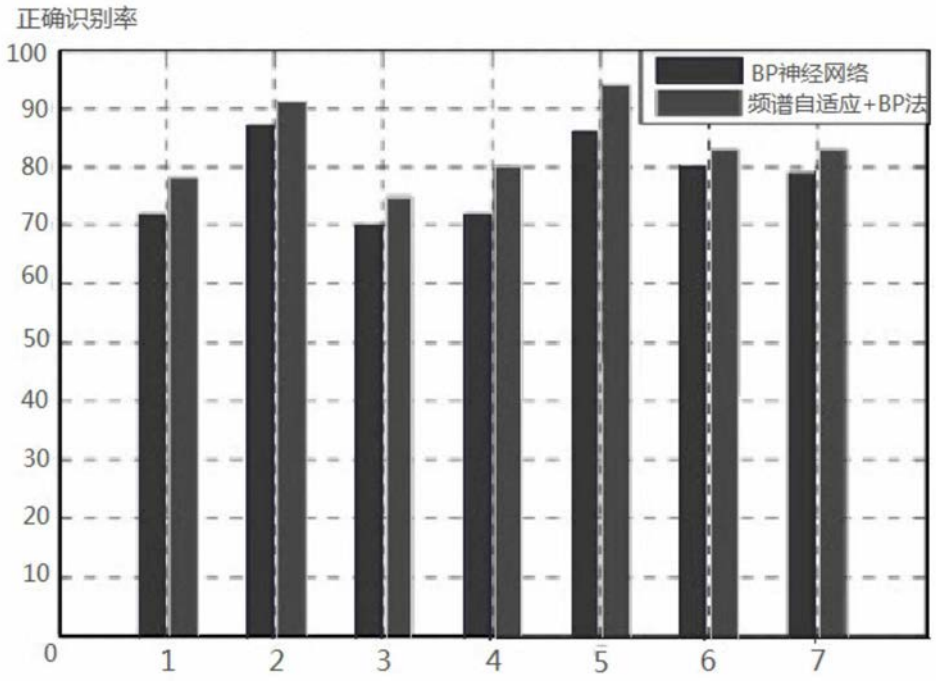


图5