



(12)发明专利申请

(10)申请公布号 CN 111583966 A

(43)申请公布日 2020.08.25

(21)申请号 202010372728.2

G10L 25/18(2013.01)

(22)申请日 2020.05.06

(71)申请人 东南大学

地址 211102 江苏省南京市江宁区东南大学路2号

(72)发明人 宗源 江林 张佳成 郑文明

江星洵 刘佳腾

(74)专利代理机构 南京苏高专利商标事务所

(普通合伙) 32204

代理人 冯艳芬

(51)Int.Cl.

G10L 25/63(2013.01)

G10L 25/27(2013.01)

G10L 25/24(2013.01)

G10L 25/21(2013.01)

权利要求书4页 说明书8页 附图1页

(54)发明名称

基于联合分布最小二乘回归的跨数据库语音情感识别方法及装置

(57)摘要

本发明公开了一种基于联合分布最小二乘回归的跨数据库语音情感识别方法及装置,方法包括:(1)获取训练数据库和测试数据库,其中,训练语音数据库中包含有若干语音片段和对应的语音情感类别标签,测试数据库中仅包含有若干待识别语音片段;(2)利用若干声学低维描述子对语音片段进行处理并进行统计,将统计得到的每个信息作为一个情感特征,并将多个情感特征组成向量作为对应语音片的特征向量;(3)建立基于联合分布的最小二乘回归模型,利用训练数据库与测试数据库联合训练,得到稀疏投影矩阵;(4)对于待识别语音片段,按照步骤(2)得到特征向量,并采用学习到的稀疏投影矩阵,得到对应的语音情感类别标签。本发明可以适应不同环境,准确率更高。

获取两个语音数据库,分别作为训练数据库和测试数据库,其中,训练语音数据库中包含有若干语音片段和对应的语音情感类别标签,测试数据库中仅包含有若干待识别语音片段

利用若干声学低维描述子对语音片段进行处理并进行统计,将统计得到的每个信息作为一个情感特征,并将多个情感特征组成向量作为对应语音片的特征向量

建立基于联合分布的最小二乘回归模型,利用已知标签的训练数据库与未知标签的测试数据库对其联合训练,得到一个连接语音片段与语音情感类别标签之间的稀疏投影矩阵

对于测试数据库中待识别语音片段,按照步骤(2)得到特征向量,并采用学习到的稀疏投影矩阵,得到对应的语音情感类别标签

1. 一种基于联合分布最小二乘回归的跨数据库语音情感识别方法,其特征在于该方法包括:

(1) 获取两个语音数据库,分别作为训练数据库和测试数据库,其中,训练语音数据库中包含有若干语音片段和对应的语音情感类别标签,测试数据库中仅包含有若干待识别语音片段;

(2) 利用若干声学低维描述子对语音片段进行处理并进行统计,将统计得到的每个信息作为一个情感特征,并将多个情感特征组成向量作为对应语音片的特征向量;

(3) 建立基于联合分布的最小二乘回归模型,利用已知标签的训练数据库与未知标签的测试数据库对其联合训练,得到一个连接语音片段与语音情感类别标签之间的稀疏投影矩阵;

(4) 对于测试数据库中待识别语音片段,按照步骤(2)得到特征向量,并采用学习到的稀疏投影矩阵,得到对应的语音情感类别标签。

2. 根据权利要求1所述的基于联合分布最小二乘回归的跨数据库语音情感识别方法,其特征在於:步骤(2)具体包括:

(2-1) 对于每个语音片段,计算其16个声学低维描述子值和对应增量参数;所述16个声学低维描述子分别为:时间信号过零率、帧能量均方根、基音频率、谐波信噪比以及梅尔顿频率倒谱系数1-12;

(2-2) 对于每个语音片段,分别对其16个声学低维描述子进行12种统计函数处理,所述12种统计函数分别为求平均值、标准差、峰态、偏度、最大值、最小值、相对位置、相对范围,以及两个线性回归系数及其均方误差;

(2-3) 将统计得到的每个信息作为一个情感特征,并将多个情感特征组成向量作为对应语音片的特征向量。

3. 根据权利要求1所述的基于联合分布最小二乘回归的跨数据库语音情感识别方法,其特征在於:步骤(3)建立的最小二乘回归模型为:

$$\min_P \left\| L_s - P^T X_s \right\|_F^2 + \lambda \left(\left\| P^T \Delta X^{st} \right\|_F^2 + \sum_{c=1}^C \left\| P^T \Delta X_c^{st} \right\|_F^2 \right) + \mu \left\| P^T \right\|_{2,1}$$

式中, $\min_P()$ 表示找到使括号内式子最小的矩阵P, $L_s \in \mathbb{R}^{c \times n}$ 为训练数据库语音片的语音情感类别标签向量, C为语音情感类别的类数, n为训练数据库语音片的个数, $X_s \in \mathbb{R}^{d \times n}$ 为训练数据库语音片的特征向量, d为特征向量的维数, $P \in \mathbb{R}^{d \times c}$ 为稀疏投影矩阵, P^T 为P的转秩矩阵, $\left\| \cdot \right\|_F^2$ 为Frobenius范数的平方, λ 、 μ 为控制正则项的权衡系数,

$\Delta X^{st} = \frac{1}{n} X_s - \frac{1}{m} X_t$, $X_t \in \mathbb{R}^{d \times m}$ 为测试数据库语音片的特征向量, m为测试数据库语音片

段的段数, $\Delta X_c^{st} = \frac{1}{n_c} \sum_{x_s \in D_s^{(c)}} X_s - \frac{1}{m_c} \sum_{x_t \in D_t^{(c)}} X_t$, $D_s^{(c)}$ 、 $D_t^{(c)}$ 分别为训练数据库、测试数据库中

情感类别属于第c类的语音片的集合, n_c 、 m_c 分别为测试数据库中情感类别属于第c类的语音片的个数, $\left\| \cdot \right\|_{2,1}$ 为2,1范数。

4. 根据权利要求3所述的基于联合分布最小二乘回归的跨数据库语音情感识别方法, 其特征在于: 步骤(3)中所述利用已知标签的训练数据库与未知标签的测试数据库对其进行联合训练的方法具体包括:

(3-1) 将所述最小二乘回归模型转换为:

$$\min_P \|L_s - Q^T X_s\|_F^2 + \lambda \left(\|Q^T \Delta X^{st}\|_F^2 + \sum_{c=1}^C \|Q^T \Delta X_c^{st}\|_F^2 \right) + \mu \|P^T\|_{2,1}$$

s. t. $P=Q$

(3-2) 通过上述转换后的最小二乘回归模型, 估算测试数据库中所有语音片段对应的语音情感类别伪标签形成的伪标签矩阵 \tilde{L}_i ;

(3-3) 根据伪标签矩阵 \tilde{L}_i 统计得到 $D_i^{(c)}$ 和 m_c , 进而计算得到 ΔX_c^{st} ;

(3-4) 基于 ΔX_c^{st} , 对转换后的最小二乘回归模型利用增广拉格朗日乘子法进行求解, 得到投影矩阵估计值 \hat{P} ;

(3-5) 根据投影矩阵估计值 \hat{P} 采用下式对伪标签矩阵 \tilde{L}_i 进行更新:

$$\begin{aligned} \bar{L}_i &= \min_{\bar{L}_i} \|L_i - \hat{P}^T X_i\|_F^2 \\ \begin{cases} \tilde{l}_i^i(k) = 1, k = \arg \max_j \bar{l}_i^i(j) \\ \tilde{l}_i^i(k) = 0, \text{其他} \end{cases} \end{aligned}$$

式中, \bar{L}_i 表示中间辅助变量, $\bar{l}_i^i(j)$ 为 \bar{L}_i 第 i 列第 j 行的元素, $\arg \max_j \bar{l}_i^i(j)$ 表示求取第 i 列元素值最大的一行的行数 j , $\tilde{l}_i^i(k)$ 是伪标签矩阵 \tilde{L}_i 第 i 列第 k 行的元素;

(3-6) 采用更新后的伪标签矩阵 \tilde{L}_i , 返回执行步骤(3-3), 直至达到预设的循环次数后, 将循环结束后得到的投影矩阵估计值 \hat{P} 作为学习得到的投影矩阵 P 。

5. 根据权利要求4所述的基于联合分布最小二乘回归的跨数据库语音情感识别方法, 其特征在于: 步骤(3-2)具体包括:

(3-2-1) 利用转换后的最小二乘回归模型不加正则项的公式, 求得投影矩阵估计值的初始值 \hat{P}_0 :

$$\hat{P}_0 = \min_P \|L_s - P^T X_s\|_F^2$$

(3-2-2) 根据投影矩阵的初始值 \hat{P}_0 采用下式得到伪标签矩阵的初始值:

$$\begin{aligned} \bar{L}_i &= \min_{\bar{L}_i} \|L_i - \hat{P}_0^T X_i\|_F^2 \\ \begin{cases} \tilde{l}_{i,0}^i(k) = 1, k = \arg \max_j \bar{l}_i^i(j) \\ \tilde{l}_{i,0}^i(k) = 0, \text{其他} \end{cases} \end{aligned}$$

式中, \bar{L}_i 表示中间辅助变量, $\tilde{L}_{i,0}^i(k)$ 是伪标签矩阵的初始值 $\tilde{L}_{i,0}$ 第 i 列第 k 行的元素。

6. 根据权利要求4所述的基于联合分布最小二乘回归的跨数据库语音情感识别方法, 其特征在于: 步骤 (3-4) 具体包括:

(3-4-1) 获取所述最小二乘回归模型的增广拉格朗日方程:

$$L(P, Q, T, \lambda, \mu) = \|L_s - Q^T X_s\|_F^2 + \lambda \left(\|Q^T \Delta X^{st}\|_F^2 + \sum_{c=1}^C \|Q^T \Delta X_c^{st}\|_F^2 \right) + \mu \|P^T\|_{2,1} \\ + \text{tr} \left[T^T (Q - P) \right] + \frac{k}{2} \|Q - P\|_F^2$$

式中, T 为拉格朗日乘子, $k > 0$ 为一个正则参数, $\text{tr}(\cdot)$ 表示求矩阵的迹;

(3-4-2) 保持 P 、 T 、 k 不变, 更新 Q :

将增广拉格朗日方程中与变量 Q 有关的部分提出, 得到:

$$\min_Q \|L_s - Q^T X_s\|_F^2 + \lambda \left(\|Q^T \Delta X^{st}\|_F^2 + \sum_{c=1}^C \|Q^T \Delta X_c^{st}\|_F^2 \right) + \text{tr} \left[T^T (Q - P) \right] + \frac{k}{2} \|Q - P\|_F^2$$

求解上式, 得到:

$$Q = [2X_s X_s^T + 2\lambda(\Delta X^{st}(\Delta X^{st})^T + \sum_{c=1}^C \Delta X_c^{st}(\Delta X_c^{st})^T + kI)]^{-1} (kP + 2X_s L_s^T - T)$$

(3-4-3) 保持 Q 、 T 、 k 不变, 更新 P :

将增广拉格朗日方程中与变量 P 有关的部分提出, 得到:

$$\min_P \mu \|P^T\|_{2,1} + \text{tr} \left[T^T (Q - P) \right] + \frac{k}{2} \|Q - P\|_F^2$$

求解上式, 得到:

$$P_i^T = \begin{cases} \frac{Q_i^T + \frac{T_i^T}{k} - \frac{\mu}{k} \left(Q_i^T + \frac{T_i^T}{k} \right)}{Q_i^T + \frac{T_i^T}{k}}, & \text{if } \left(\frac{\mu}{k} < Q_i^T + \frac{T_i^T}{k} \right) \\ 0, & \text{其他} \end{cases}$$

P_i 是 P 的第 i 个列向量, T_i 是 T 的第 i 个列向量;

(3-4-4) 保持 Q 、 P 不变, 更新 T 、 k :

$$T = T + k(P - C)$$

$$k = \min(\rho k, k_{\max})$$

式中, k_{\max} 是预设 k 的最大值, ρ 是缩放系数, $\rho > 1$;

(3-4-5) 检查是否收敛:

检查 $\|P - Q\|_{\infty} < \varepsilon$ 是否成立, 若否, 则返回执行步骤 (3-4-2), 若是或迭代次数大于设置值, 则将此时的 P 的值作为所求的稀疏投影矩阵, $\|\cdot\|_{\infty}$ 表示求数据中的最大元素, ε 表示收敛阈值。

7. 根据权利要求1所述的基于联合分布最小二乘回归的跨数据库语音情感识别方法, 其特征在于: 步骤 (4) 中所述测试数据库语音情感类别标签的计算方法为:

采用下式计算:

$$\bar{L}_t = \min_{L_t} \|L_t - P^T X_t\|_F^2$$
$$j^* = \arg \max_j \bar{L}_t^i(j)$$

式中,P为步骤(3)学习到的投影矩阵, X_t 表示测试数据库语音片的特征向量集合,即待识别语音片的特征向量集合, \bar{L}_t 表示中间辅助变量, j^* 表示待识别语音片的语音情感类别标签。

8.一种基于联合分布最小二乘回归的跨数据库语音情感识别装置,包括处理器及存储在存储器上并可在处理器上运行的计算机程序,其特征在于:所述处理器执行所述程序时实现权利要求1-6中任意一项所述的方法。

基于联合分布最小二乘回归的跨数据库语音情感识别方法及装置

技术领域

[0001] 本发明涉及语音情感识别,尤其涉及一种基于联合分布最小二乘回归的跨数据库语音情感识别方法及装置。

背景技术

[0002] 语音情感识别的目的在于使得机器能够拥有足够智能从说话者的语音中提取它的情感状态(如高兴、恐惧、悲伤等),是人机交互中重要的一个环节,拥有巨大的研究潜能与发展前景。如结合驾驶员的语音、表情和行为信息检测其精神状态,可以及时提醒驾驶员集中注意力避免危险驾驶;在人机交互中检测对话人的语音情感可以使得对话更加流畅,更加照顾对话者的心理,贴近认知;可穿戴设备可以依据穿戴者的情感状态做出更为及时和贴心的反馈;同时,在课堂教学、老师陪护等各种各样的领域,语音情感识别都在发挥着越来越重要的作用。

[0003] 传统的语音情感识别都在同一个语音数据库上进行训练和测试,训练和测试的数据都遵循着同样的分布。而在实际生活中,训练出的模型必须面对不同的环境,发声背景中也会掺杂着各种各样的噪音。因此跨数据库语音情感识别面临着很大的挑战。如何使训练出的模型面对不同的环境都拥有良好的适应性,成为学术和工业界需要解决的问题。

发明内容

[0004] 发明目的:本发明针对现有技术存在的问题,提供一种基于联合分布最小二乘回归的跨数据库语音情感识别方法及装置,本发明对于不同环境都拥有良好的适应性,识别结果更准确。

[0005] 技术方案:本发明所述的基于联合分布最小二乘回归的跨数据库语音情感识别方法包括:

[0006] (1) 获取两个语音数据库,分别作为训练数据库和测试数据库,其中,训练语音数据库中包含有若干语音片段和对应的语音情感类别标签,而测试数据库中仅包含有若干待识别语音片段;

[0007] (2) 利用若干声学低维描述子对语音片段进行处理并进行统计,将统计得到的每个信息作为一个情感特征,并将多个情感特征组成向量作为对应语音片的特征向量;

[0008] (3) 建立基于联合分布的最小二乘回归模型,利用已知标签的训练数据库与未知标签的测试数据库对其联合训练,得到一个连接语音片段与语音情感类别标签之间的稀疏投影矩阵;

[0009] (4) 对于测试数据库中待识别语音片段,按照步骤(2)得到特征向量,并采用学习到的稀疏投影矩阵,得到对应的语音情感类别标签。

[0010] 进一步的,步骤(2)具体包括:

[0011] (2-1) 对于每个语音片段,计算其16个声学低维描述子值和对应增量参数;所述16

个声学低维描述子分别为:时间信号过零率、帧能量均方根、基音频率、谐波信噪比以及梅尔顿频率倒谱系数1-12;

[0012] (2-2) 对于每个语音片段,分别对其16个声学低维描述子进行12种统计函数处理,所述12种统计函数分别为求平均值、标准差、峰态、偏度、最大值、最小值、相对位置、相对范围,以及两个线性回归系数及其均方误差;

[0013] (2-3) 将统计得到的每个信息作为一个情感特征,并将多个情感特征组成向量作为对应语音片的特征向量。

[0014] 进一步的,步骤(3)建立的最小二乘回归模型为:

$$[0015] \quad \min_P \left\| L_s - P^T X_s \right\|_F^2 + \lambda \left(\left\| P^T \Delta X^{st} \right\|_F^2 + \sum_{c=1}^C \left\| P^T \Delta X_c^{st} \right\|_F^2 \right) + \mu \left\| P^T \right\|_{2,1}$$

[0016] 式中, $\min_P()$ 表示找到使括号内式子最小的矩阵P, $L_s \in \mathbb{R}^{c \times n}$ 为训练数据库语音片段的语音情感类别标签向量, C为语音情感类别的类数, n为训练数据库语音片段的个数, $X_s \in \mathbb{R}^{d \times n}$ 为训练数据库语音片的特征向量, d为特征向量的维数, $P \in \mathbb{R}^{d \times c}$ 为稀疏投影矩阵, P^T 为P的转秩矩阵, $\left\| \cdot \right\|_F^2$ 为Frobenius范数的平方, λ 、 μ 为控制正则项的权衡系数,

$\Delta X^{st} = \frac{1}{n} X_s - \frac{1}{m} X_t$, $X_t \in \mathbb{R}^{d \times m}$ 为测试数据库语音片的特征向量, m为测试数据库语音

片的段数, $\Delta X_c^{st} = \frac{1}{n_c} \sum_{x_s \in D_s^{(c)}} X_s - \frac{1}{m_c} \sum_{x_t \in D_t^{(c)}} X_t$, $D_s^{(c)}$ 、 $D_t^{(c)}$ 分别为训练数据库、测试数据库

中情感类别属于第c类的语音片段的集合, n_c 、 m_c 分别为测试数据库中情感类别属于第c类的语音片的个数, $\left\| \cdot \right\|_{2,1}$ 为2,1范数。

[0017] 进一步的,步骤(3)中所述利用已知标签的训练数据库与未知标签的测试数据库对其进行联合训练的方法具体包括:

[0018] (3-1) 将所述最小二乘回归模型转换为:

$$[0019] \quad \min_P \left\| L_s - Q^T X_s \right\|_F^2 + \lambda \left(\left\| Q^T \Delta X^{st} \right\|_F^2 + \sum_{c=1}^C \left\| Q^T \Delta X_c^{st} \right\|_F^2 \right) + \mu \left\| P^T \right\|_{2,1}$$

[0020] s. t. $P=Q$

[0021] (3-2) 通过上述转换后的最小二乘回归模型,估算测试数据库中所有语音片对应的语音情感类别伪标签形成的伪标签矩阵 \tilde{L}_t ;

[0022] (3-3) 根据伪标签矩阵 \tilde{L}_t 统计得到 $D_t^{(c)}$ 和 m_c , 进而计算得到 ΔX_c^{st} ;

[0023] (3-4) 基于 ΔX_c^{st} , 对转换后的最小二乘回归模型利用增广拉格朗日乘子法进行求解, 得到投影矩阵估计值 \hat{P} ;

[0024] (3-5) 根据投影矩阵估计值 \hat{P} 采用下式对伪标签矩阵 \tilde{L}_t 进行更新:

$$[0025] \quad \bar{L}_t = \min_{L_t} \left\| L_t - \hat{P}^T X_t \right\|_F^2$$

$$[0026] \quad \begin{cases} \tilde{l}_t^i(k) = 1, k = \arg \max_j \bar{l}_t^i(j) \\ \tilde{l}_t^i(k)=0, \text{其他} \end{cases}$$

[0027] 式中, \bar{L}_t 表示中间辅助变量, $\bar{l}_t^i(j)$ 为 \bar{L}_t 第 i 列第 j 行的元素, $\arg \max_j \bar{l}_t^i(j)$ 表示求取第 i 列元素值最大的一行的行数 j, $\tilde{l}_t^i(k)$ 是伪标签矩阵 \tilde{L}_t 第 i 列第 k 行的元素;

[0028] (3-6) 采用更新后的伪标签矩阵 \tilde{L}_t , 返回执行步骤 (3-3), 直至达到预设的循环次数后, 将循环结束后得到的投影矩阵估计值 \hat{P} 作为学习得到的投影矩阵 P。

[0029] 进一步的, 步骤 (3-2) 具体包括:

[0030] (3-2-1) 利用转换后的最小二乘回归模型不加正则项的公式, 求得投影矩阵估计值的初始值 \hat{P}_0 :

$$[0031] \quad \hat{P}_0 = \min_P \|L_s - P^T X_s\|_F^2$$

[0032] (3-2-2) 根据投影矩阵的初始值 \hat{P}_0 采用下式得到伪标签矩阵的初始值:

$$[0033] \quad \bar{L}_t = \min_{\bar{L}_t} \|L_t - \hat{P}_0^T X_t\|_F^2$$

$$[0034] \quad \begin{cases} \tilde{l}_{t,0}^i(k) = 1, k = \arg \max_j \bar{l}_t^i(j) \\ \tilde{l}_{t,0}^i(k)=0, \text{其他} \end{cases}$$

[0035] 式中, \bar{L}_t 表示中间辅助变量, $\tilde{l}_{t,0}^i(k)$ 是伪标签矩阵的初始值 $\tilde{L}_{t,0}$ 第 i 列第 k 行的元素。进一步的, 步骤 (3-4) 具体包括:

[0036] (3-4-1) 获取所述最小二乘回归模型的增广拉格朗日方程:

$$[0037] \quad L(P, Q, T, \lambda, \mu) = \|L_s - Q^T X_s\|_F^2 + \lambda \left(\|Q^T \Delta X^{st}\|_F^2 + \sum_{c=1}^C \|Q^T \Delta X_c^{st}\|_F^2 \right) + \mu \|P^T\|_{2,1} \\ + \text{tr} [T^T (Q - P)] + \frac{k}{2} \|Q - P\|_F^2$$

[0038] 式中, T 为拉格朗日乘子, k > 0 为一个正则参数, tr() 表示求矩阵的迹;

[0039] (3-4-2) 保持 P、T、k 不变, 更新 Q:

[0040] 将增广拉格朗日方程中与变量 Q 有关的部分提出, 得到:

$$[0041] \quad \min_Q \|L_s - Q^T X_s\|_F^2 + \lambda \left(\|Q^T \Delta X^{st}\|_F^2 + \sum_{c=1}^C \|Q^T \Delta X_c^{st}\|_F^2 \right) + \text{tr} [T^T (Q - P)] + \frac{k}{2} \|Q - P\|_F^2$$

[0042] 求解上式, 得到:

$$[0043] \quad Q = [2X_s X_s^T + 2\lambda(\Delta X^{st}(\Delta X^{st})^T + \sum_{c=1}^C \Delta X_c^{st}(\Delta X_c^{st})^T + kI)]^{-1} (kP + 2X_s L_s^T - T)$$

[0044] (3-4-3) 保持 Q、T、k 不变, 更新 P:

[0045] 将增广拉格朗日方程中与变量 P 有关的部分提出, 得到:

$$[0046] \quad \min_P \mu \left\| P^T \right\|_{2,1} + \text{tr} \left[T^T (Q - P) \right] + \frac{k}{2} \|Q - P\|_F^2$$

[0047] 求解上式,得到:

$$[0048] \quad P_i^T = \begin{cases} \frac{Q_i^T + \frac{T_i^T}{k} - \frac{\mu}{k} \left(Q_i^T + \frac{T_i^T}{k} \right)}{Q_i^T + \frac{T_i^T}{k}}, & \text{if } \left(\frac{\mu}{k} < Q_i^T + \frac{T_i^T}{k} \right) \\ 0, & \text{其他} \end{cases}$$

[0049] P_i 是 P 的第 i 个列向量, T_i 是 T 的第 i 个列向量;

[0050] (3-4-4) 保持 Q 、 P 不变,更新 T 、 k :

$$[0051] \quad T = T + k(P - C)$$

$$[0052] \quad k = \min(\rho k, k_{\max})$$

[0053] 式中, k_{\max} 是预设 k 的最大值, ρ 是缩放系数, $\rho > 1$;

[0054] (3-4-5) 检查是否收敛:

[0055] 检查 $\|P - Q\|_{\infty} < \varepsilon$ 是否成立,若否,则返回执行步骤(3-4-2),若是或迭代次数大于设置值,则将此时的 P 的值作为所求的稀疏投影矩阵, $\|\cdot\|_{\infty}$ 表示求数据中的最大元素, ε 表示收敛阈值。

[0056] 进一步的,步骤(4)中所述测试数据库的语音情感类别标签的计算方法为:

[0057] 采用下式计算:

$$[0058] \quad \bar{L}_i = \min_{L_i} \|L_i - P^T X_i\|_F^2$$

$$[0059] \quad j^* = \arg \max_j \bar{L}_i^j(j)$$

[0060] 式中, P 为我们学习到的最终的投影矩阵, X_i 表示测试数据库语音片的特征向量集合,即待识别语音片的特征向量集合, \bar{L}_i 表示中间辅助变量, j^* 表示待识别语音片的语音情感类别标签。

[0061] 本发明所述的基于联合分布最小二乘回归的跨数据库语音情感识别装置包括处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现上述方法。

[0062] 有益效果:本发明与现有技术相比,其显著优点是:本发明的跨数据库语音情感识别方法及装置是在跨库学习,因此,对于不同环境都拥有良好的适应性,识别结果更准确。

附图说明

[0063] 图1是本发明提供的基于联合分布最小二乘回归的跨数据库语音情感识别方法的流程示意图。

具体实施方式

[0064] 本实施例提供了一种基于联合分布最小二乘回归的跨数据库语音情感识别方法,如图1所示,包括以下步骤:

[0065] (1) 获取两个语音数据库,分别作为训练数据库和测试数据库,其中,训练语音数据库中包有若干语音片段和对应的语音情感类别标签,而测试数据库中仅包有若干待识别语音片段。

[0066] 在本实施例中,我们使用了情感语音识别中常见的三类语音情感数据库:Berlin、eINTERFACE和CAISA。因为三类数据库包含的情感类别不同,所以在两两比较时都对数据进行了选择。当Berlin和eINTERFACE进行比较时,我们分别选取了375条数据和1077条数据,情感类别为5类(生气、害怕、快乐、厌恶、悲伤);当Berlin和CAISA进行比较时,我们分别选取了408条数据和1000条数据,情感类别为5类(生气、害怕、高兴、厌恶、悲伤);当eINTERFACE和CAISA进行比较时,我们分别选取了1072条数据和1000条数据,情感类别为5类(生气、害怕、高兴、厌恶、悲伤)。

[0067] (2) 利用若干声学低维描述子对语音片段进行处理并进行统计,将统计得到的每个信息作为一个情感特征,并将多个情感特征组成向量作为对应语音片的特征向量。

[0068] 该步骤具体包括:

[0069] (2-1) 对于每个语音片段,计算其16个声学低维描述子值和对应增量参数;所述16个声学低维描述子分别为:时间信号过零率、帧能量均方根、基音频率、谐波信噪比以及梅尔顿频率倒谱系数1-12;描述子来源于INTERSPEECH 2009 Emotion Challenge提供的功能集;

[0070] (2-2) 对于每个语音片段,利用openSIMLE软件分别对其16个声学低维描述子进行12种统计函数处理,所述12种统计函数分别为求平均值、标准差、峰态、偏度、最大值、最小值、相对位置、相对范围,以及两个线性回归系数及其均方误差;

[0071] (2-3) 将统计得到的每个信息作为一个情感特征,并将 $16 \times 2 \times 12 = 384$ 个情感特征组成向量作为对应语音片的特征向量。

[0072] (3) 建立基于联合分布的最小二乘回归模型,利用已知标签的训练数据库与未知标签的测试数据库对其联合训练,得到一个连接语音片段与语音情感类别标签之间的稀疏投影矩阵。

[0073] 其中,建立的最小二乘回归模型为:

$$[0074] \quad \min_P \left\| L_s - P^T X_s \right\|_F^2 + \lambda \left(\left\| P^T \Delta X^{st} \right\|_F^2 + \sum_{c=1}^C \left\| P^T \Delta X_c^{st} \right\|_F^2 \right) + \mu \left\| P^T \right\|_{2,1}$$

[0075] 式中, $\min_P()$ 表示找到使括号内式子最小的矩阵P, $L_s \in \mathbb{R}^{c \times n}$ 为训练数据库语音片段的语音情感类别标签向量, C为语音情感类别的类数, n为训练数据库语音片段的个数, $X_s \in \mathbb{R}^{d \times n}$ 为训练数据库语音片的特征向量, d为特征向量的维数, $P \in \mathbb{R}^{d \times c}$ 为稀疏投影矩阵, P^T 为P的转秩矩阵, $\left\| \cdot \right\|_F^2$ 为Frobenius范数的平方, λ 、 μ 为控制正则项的权衡系数,

$\Delta X^{st} = \frac{1}{n} X_s - \frac{1}{m} X_t$, $X_t \in \mathbb{R}^{d \times m}$ 为测试数据库语音片的特征向量, m为测试数据库语音片

段的段数, $\Delta X_c^{st} = \frac{1}{n_c} \sum_{x_s \in D_s^{(c)}} X_s - \frac{1}{m_c} \sum_{x_t \in D_t^{(c)}} X_t$, $D_s^{(c)}$ 、 $D_t^{(c)}$ 分别为训练数据库、测试数据库中

情感类别属于第c类的语音片段的集合, n_c 、 m_c 分别为测试数据库中情感类别属于第c类的语

音片段的个数, $\| \cdot \|_{2,1}$ 为 $2,1$ 范数。

[0076] 其中, 利用已知标签的训练数据库与未知标签的测试数据库对其联合训练的方法具体包括:

[0077] (3-1) 将所述最小二乘回归模型转换为:

$$[0078] \quad \min_P \left\| L_s - Q^T X_s \right\|_F^2 + \lambda \left(\left\| Q^T \Delta X^{st} \right\|_F^2 + \sum_{c=1}^C \left\| Q^T \Delta X_c^{st} \right\|_F^2 \right) + \mu \left\| P^T \right\|_{2,1}$$

[0079] s. t. $P=Q$

[0080] (3-2) 通过上述转换后的最小二乘回归模型, 估算测试数据库中所有语音片段对应的语音情感类别伪标签形成的伪标签矩阵 \tilde{L}_i ;

[0081] (3-3) 根据伪标签矩阵 \tilde{L}_i 统计得到 $D_i^{(c)}$ 和 m_c , 进而计算得到 ΔX_c^{st} ;

[0082] (3-4) 基于 ΔX_c^{st} , 对转换后的最小二乘回归模型利用增广拉格朗日乘子法进行求解, 得到投影矩阵估计值 \hat{P} ;

[0083] (3-5) 根据投影矩阵估计值 \hat{P} 采用下式对伪标签矩阵 \tilde{L}_i 进行更新:

$$[0084] \quad \bar{L}_i = \min_{L_i} \left\| L_i - \hat{P}^T X_i \right\|_F^2$$

$$[0085] \quad \begin{cases} \tilde{l}_i^i(k) = 1, k = \arg \max_j \bar{l}_i^i(j) \\ \tilde{l}_i^i(k) = 0, \text{ 其他} \end{cases}$$

[0086] 式中, \bar{L}_i 表示中间辅助变量, $\bar{l}_i^i(j)$ 为 \bar{L}_i 第 i 列第 j 行的元素, $\arg \max_j \bar{l}_i^i(j)$ 表示求取第 i 列元素值最大的一行的行数 j , $\tilde{l}_i^i(k)$ 是伪标签矩阵 \tilde{L}_i 第 i 列第 k 行的元素;

[0087] (3-6) 采用更新后的伪标签矩阵 \tilde{L}_i , 返回执行步骤 (3-3), 直至达到预设的循环次数后, 将循环结束后得到的投影矩阵估计值 \hat{P} 作为学习得到的投影矩阵 P 。

[0088] 进一步的, 步骤 (3-2) 具体包括:

[0089] (3-2-1) 利用转换后的最小二乘回归模型不加正则项的公式, 求得投影矩阵估计值的初始值 \hat{P}_0 :

$$[0090] \quad \hat{P}_0 = \min_P \left\| L_s - P^T X_s \right\|_F^2$$

[0091] (3-2-2) 根据投影矩阵的初始值 \hat{P}_0 采用下式得到伪标签矩阵的初始值:

$$[0092] \quad \bar{L}_i = \min_{L_i} \left\| L_i - \hat{P}_0^T X_i \right\|_F^2$$

$$[0093] \quad \begin{cases} \tilde{l}_{i,0}^i(k) = 1, k = \arg \max_j \bar{l}_i^i(j) \\ \tilde{l}_{i,0}^i(k) = 0, \text{ 其他} \end{cases}$$

[0094] 式中, \bar{L}_i 表示中间辅助变量, $\tilde{l}_{i,0}^i(k)$ 是伪标签矩阵的初始值 $\tilde{L}_{i,0}$ 第 i 列第 k 行的元素。

伪标签矩阵 L_c 的每一列只有其对应的类别那一行为1,其余行都为0。

[0095] 步骤(3-4)具体包括:

[0096] (3-4-1)获取所述最小二乘回归模型的增广拉格朗日方程:

$$L(P, Q, T, \lambda, \mu) = \|L_s - Q^T X_s\|_F^2 + \lambda \left(\|Q^T \Delta X^{st}\|_F^2 + \sum_{c=1}^C \|Q^T \Delta X_c^{st}\|_F^2 \right) + \mu \|P^T\|_{2,1} \\ + \text{tr} [T^T (Q - P)] + \frac{k}{2} \|Q - P\|_F^2$$

[0098] 式中,T为拉格朗日乘子,k>0为一个正则参数, tr()表示求矩阵的迹;

[0099] (3-4-2)保持P、T、k不变,更新Q:

[0100] 将增广拉格朗日方程中与变量Q有关的部分提出,得到:

$$\min_Q \|L_s - Q^T X_s\|_F^2 + \lambda \left(\|Q^T \Delta X^{st}\|_F^2 + \sum_{c=1}^C \|Q^T \Delta X_c^{st}\|_F^2 \right) + \text{tr} [T^T (Q - P)] + \frac{k}{2} \|Q - P\|_F^2$$

[0102] 求解上式,得到:

$$Q = [2X_s X_s^T + 2\lambda(\Delta X^{st}(\Delta X^{st})^T + \sum_{c=1}^C \Delta X_c^{st}(\Delta X_c^{st})^T + kI)]^{-1} (kP + 2X_s L_s^T - T)$$

[0104] (3-4-3)保持Q、T、k不变,更新P:

[0105] 将增广拉格朗日方程中与变量P有关的部分提出,得到:

$$\min_P \mu \|P^T\|_{2,1} + \text{tr} [T^T (Q - P)] + \frac{k}{2} \|Q - P\|_F^2$$

[0107] 求解上式,得到:

$$P_i^T = \begin{cases} \frac{Q_i^T + \frac{T_i^T}{k} - \frac{\mu}{k} \left(Q_i^T + \frac{T_i^T}{k} \right)}{Q_i^T + \frac{T_i^T}{k}}, & \text{if } \left(\frac{\mu}{k} < Q_i^T + \frac{T_i^T}{k} \right) \\ 0, & \text{其他} \end{cases}$$

[0109] P_i 是P的第i个列向量, T_i 是T的第i个列向量;

[0110] (3-4-4)保持Q、P不变,更新T、k:

$$T = T + k(P - C)$$

$$k = \min(\rho k, k_{\max})$$

[0113] 式中, k_{\max} 是预设k的最大值, ρ 是缩放系数, $\rho > 1$;

[0114] (3-4-5)检查是否收敛:

[0115] 检查 $\|P - Q\|_{\infty} < \varepsilon$ 是否成立,若否,则返回执行步骤(3-4-2),若是或迭代次数大于设置值,则将此时的P的值作为所求的稀疏投影矩阵, $\|\cdot\|_{\infty}$ 表示求数据中的最大元素,

[0116] ε 表示收敛阈值。

[0117] (4)对于测试数据库中待识别语音片段,按照步骤(2)得到特征向量,并采用学习到的稀疏投影矩阵,得到对应的语音情感类别标签。

[0118] 具体方法为采用下式计算类别标签:

$$[0119] \quad \bar{L}_i = \min_{L_i} \|L_i - P^T X_i\|_F^2$$

$$[0120] \quad j^* = \arg \max_j \bar{L}_i^i(j)$$

[0121] 式中,P为我们学习到的最终的投影矩阵, X_i 表示测试数据库语音片的特征向量集合,即待识别语音片的特征向量集合, \bar{L}_i 表示中间辅助变量, j^* 表示待识别语音片的语音情感类别标签。

[0122] 本实施例还提供了一种基于联合分布最小二乘回归的跨数据库语音情感识别装置,包括处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现上述方法。

[0123] 为了验证本发明的有效性,在语音情感数据库Berlin、eNTERFACE和CAISA数据库上我们两两分别进行了实验。在每一组实验中,我们将两种数据库分别作为源域和目标域,其中源域是作为训练集提供训练数据和标签,目标域是作为测试集,仅提供测试数据,不提供任何标签。为了更有效的检测识别准确率,我们采用了非加权平均召回率(UAR)与加权平均召回率(WAR)两种检测方法。其中,UAR表示每一类正确预测的数量除以每一类参与测试的数量,再对所有类的正确率取代数平均;而WAR表示所有正确预测的数量除以所有参与测试的数量,不考虑每一类数量的影响。综合考虑UAR和 WAR可以有效避免类别数量不平衡的影响。作为对比实验,我们选取了子空间学习中经典且高效的几类算法,分别为:SVM、TCA、TKL、DaLSR、DoSL。实验结果如下表1所示,其中,本方法在表中表示为英文缩写JDLSR,数据集为源域/目标域,E、B、C分别为eNTERFACE、Berlin、CASIA数据集的缩写,评价标准为UAR/WAR。

[0124] 实验结果表明,基于本发明提出的微表情识别方法,取得了较高的跨数据库微表情识别率。

[0125] 表1

[0126]

数据集 实验	E/B	B/E	B/C	C/B	E/C	C/E
SVM	27.83/24.27	30.06/30.08	25.10/25.10	33.59/36.76	24.20/24.20	26.26/26.31
TCA	29.70/39.20	26.25/26.28	28.10/28.10	37.24/37.99	26.10/26.10	24.20/24.25
TKL	36.21/42.40	24.55/24.61	25.10/25.10	35.08/38.24	24.10/24.10	27.48/27.52
DaLSR	44.41/52.27	36.36/36.40	25.40/25.40	23.85/26.96	20.00/20.00	22.65/24.21
DoSL	44.25/52.00	37.49/37.51	34.40/34.40	42.40/45.82	22.70/18.92	25.56/21.29
JDLSR	44.49/52.27	37.48/37.51	33.30/33.30	47.74/46.08	27.10/27.10	27.85/27.89

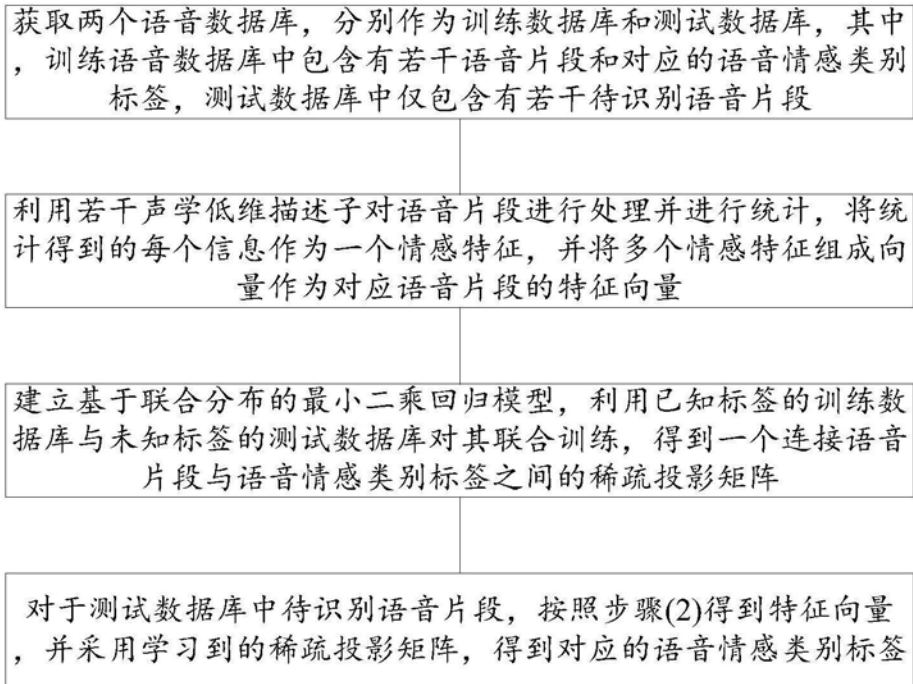


图1