



面向数据增强的多种语音情感分类算法研究

李茜茜, 沈晓燕, 任福继, 康鑫

引用本文:

李茜茜, 沈晓燕, 任福继, 等. 面向数据增强的多种语音情感分类算法研究[J]. 智能系统学报, 2021, 16(1): 170–177.

LI Qianqian, SHEN Xiaoyan, REN Fuji, et al. Investigation of multiple speech emotion classification algorithms based on data enhancement[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(1): 170–177.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202103005>

您可能感兴趣的其他文章

语音情感识别研究综述

Review on speech emotion recognition research

智能系统学报. 2020, 15(1): 1–13 <https://dx.doi.org/10.11992/tis.201904065>

触觉手势情感识别的超限学习方法

Extreme learning machine for emotion recognition of tactile gestures

智能系统学报. 2019, 14(1): 127–133 <https://dx.doi.org/10.11992/tis.201804029>

基于测度学习支持向量机的钢琴乐谱难度等级识别

Recognition of difficulty level of piano score based on metric learning support vector machine

智能系统学报. 2018, 13(2): 196–201 <https://dx.doi.org/10.11992/tis.201612012>

维度加权模式动态纹理特征的火焰检测

Fire detection based on dynamic texture features under a dimension-weighted mode

智能系统学报. 2017, 12(4): 548–555 <https://dx.doi.org/10.11992/tis.201607021>

多特征的光学遥感图像多目标识别算法

Research on multi-feature based multi-target recognition algorithm for optical remote sensing image

智能系统学报. 2016, 11(5): 655–662 <https://dx.doi.org/10.11992/tis.201511011>

智能手机车辆异常驾驶行为检测方法

Abnormal driving behavior detection based on the smart phone

智能系统学报. 2016, 11(3): 410–417 <https://dx.doi.org/10.11992/tis.201504022>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202103005

面向数据增强的多种语音情感分类算法研究

李茜茜^{1,2}, 沈晓燕¹, 任福继², 康鑫²

(1. 南通大学 信息科学技术学院, 江苏 南通 226019; 2. 日本德岛大学 智能信息工学部, 日本 德岛 7708501)

摘 要: 目前语音情感识别存在语音样本不足、提取的特征数据量大和无关特征多使得识别率不高的问题。针对语音样本不足的情况, 在预处理阶段提出了时频域的数据增强方法, 对原有的数据库进行扩充; 根据传统算法中提取的特征数据量大且与情感无关的特征多的现状, 提取了 1 582 维的情感特征和 10 组低级描述特征。分别在支持向量机、随机森林和 K 最邻近 3 种机器学习算法上做了对比实验。实验证明: 支持向量机的平均识别率比较好。在所提取的 10 组特征组中, LogMelFreqBand 特征在 3 种算法上的精确度分别为 74.63%、64.93% 和 66.42%; 而 pcm_fftMag_mfcc 特征的精确度分别为 84.33%、73.13% 和 58.21%。

关键词: 语音情感识别; 数据增强; 情感特征; 支持向量机; 随机森林; K 最邻近; 低级描述特征; 机器学习
中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2021)01-0170-08

中文引用格式: 李茜茜, 沈晓燕, 任福继, 等. 面向数据增强的多种语音情感分类算法研究 [J]. 智能系统学报, 2021, 16(1): 170-177.

英文引用格式: LI Qianqian, SHEN Xiaoyan, REN Fuji, et al. Investigation of multiple speech emotion classification algorithms based on data enhancement[J]. CAAI transactions on intelligent systems, 2021, 16(1): 170-177.

Investigation of multiple speech emotion classification algorithms based on data enhancement

LI Qianqian^{1,2}, SHEN Xiaoyan¹, REN Fuji², KANG Xin²

(1. Institute of Information Science and Technology, Nantong University, Nantong 226019, China; 2. Department of Intelligent Information Engineering, Tokushima University, Tokushima 7708501, Japan)

Abstract: Currently, problems in speech emotion recognition, such as insufficient speech samples and numerous extracted and irrelevant features, make the recognition rate low. To solve the problem of insufficient speech samples, a time-frequency domain data enhancement method is proposed in the preprocessing stage to expand the original database. Considering the current situation where traditional algorithms extract a large amount of feature data and many are emotion-independent, 1582-dimensional emotion features and 10 groups of low-level description features were extracted. Finally, a comparative experiment was performed on three machine learning algorithms: the support vector machine, random forest, and K -nearest neighbor. Experiments showed that the average recognition rate of the support vector machine was superior. Among the ten sets of features, the accuracy of LogMelFreqBand in the three algorithms was 74.63%, 64.93%, and 66.42%, respectively, and the accuracy of pcm_fftMag_mfcc was 84.33%, 73.13%, and 58.21%, respectively.

Keywords: speech emotion recognition; data enhancement; emotion feature; support vector machine; random forest; K -nearest neighbor; low-level description features; machine learning

语音情感识别是人工智能的一个重大研究领域。人机交互中, 语音是最直接的交流方式, 对人的情感和认知的研究是人工智能的一个质的进步。语音情感识别是利用计算机对输入语音的情绪状态进行自动化识别。它在心理健康监测、教

育辅助、个性化内容推荐、客户服务质量监测等领域具有广阔的应用前景。

目前的语音情感识别主要由以下几个过程构成: 预处理、特征提取和情感分类。预处理方面的研究主要集中在数据库的处理和降噪等方面。当使用机器学习进行语音情感识别时, 经常需要大量的训练数据, 若数据集太少, 则会出现过拟合的缺陷。为了能够在小样本数据集上进行算法

收稿日期: 2021-03-16.

基金项目: 国家自然科学基金项目 (61534003, 81371663); 德岛大学研究集群项目 (2003002).

通信作者: 沈晓燕. E-mail: xiaoyansho@ntu.edu.cn.

的训练识别, 对样本引进增强变换, 增加样本数据量^[1], 可以达到提高识别率的目的。Chatziagapi 等^[2]使用生成对抗性网络 (generative adversarial nets, GANs) 生成声谱图, 对原有的样本进行数据增强, 在 IEMOCAP 和 FEEL-25k 情感数据库上得到的未加权平均召回率为 53.6% 和 54.6%。语音样本中夹杂着混响和附加噪声, 这对语音情感识别的性能有不利影响。在预处理阶段, 减少或消除混响和加性噪声称为去噪, 也称为语音增强^[3]。任福继等^[4]在 2016 年提出了一种通过人工修改标签和降低噪声来对数据集进行预处理的方法。川濑友子等^[5]提出了一种自动切换降噪参数的方法, 使得自动语音识别的准确率最大化。然而, 大多数的去噪方法都不能很好地滤除噪声, 因为信噪比和峰值信噪比太小。由于不可能对这种高噪声数据进行完全去噪, Yousefi 等^[6]提出采用增强方法来提高数据质量。但是语音情感识别领域中, 使用数据增强方法来提高识别率的研究较少, 多被用于图像识别。

与特征提取方面相关的研究有特征选择和特征融合等。Reda Elbarougy 等^[7]认为, 在提取情感特征时, 应该考虑声学特征和情感状态之间的相关性。在语音情感识别中, 特征选择是提高识别

准确率的重要步骤。从语音信号中可以提取出多个特征, 然而哪一组特征与说话人的情绪状态最相关还在研究中。到目前为止, 与情绪状态相关的特征少之又少。关于情感识别分类阶段, 也有不少研究者提出了各种情感分类算法, 并推陈出新。Karttikeya 等^[8]提出了 2 种监督式学习的语音情感识别方法, 但是由于训练样本稀少, 在 IEMOCAP 语料库的识别率只达到了 69.1%。

本文针对语音样本不足和现阶段语音情感识别提取的无关特征多的问题, 提出了面向数据增强的语音情感识别方法。

1 研究方法

本节介绍了一种面向数据增强的语音情感识别方法, 其系统总体结构如图 1 所示。系统包括预处理、特征提取与分类识别 3 个过程。在预处理阶段, 原始数据集的声音样本首先经过时域和频域 2 个方向的数据增强形成新的样本。新生成的样本与原始的情感数据库被一起送入特征提取部分。在特征提取过程, 使用 openSMILE2.3.0 提取 1 582 维的情感特征和 10 组低级描述特征, 然后按照训练和测试比例 6:1 的概率送入 3 种分类算法模型中, 最后对比得到每种情感的识别结果。

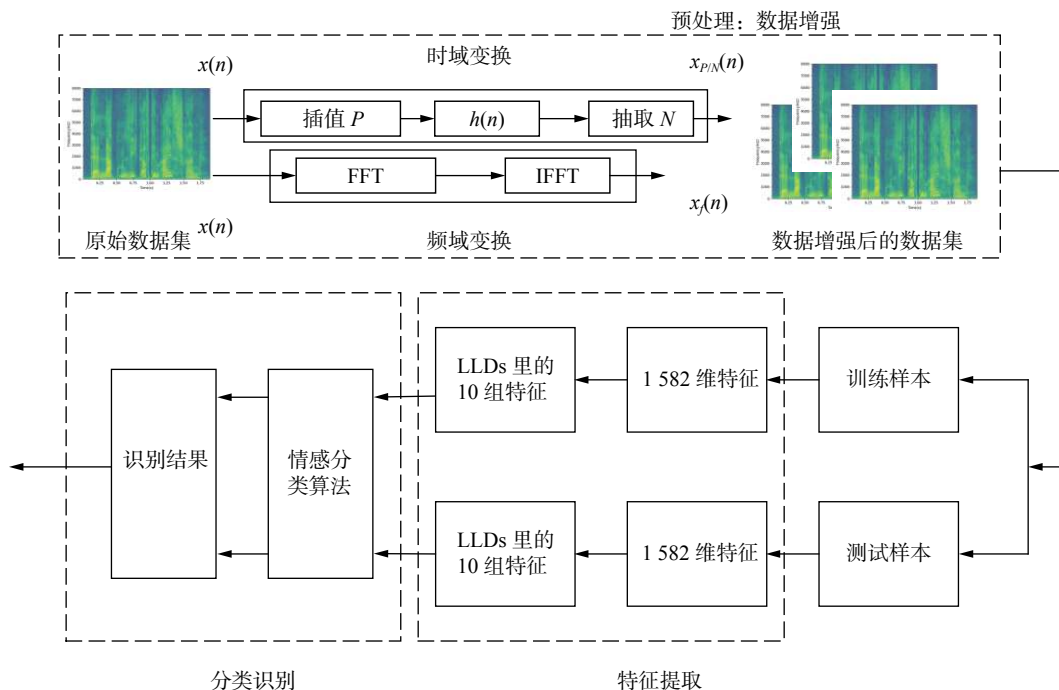


图 1 基于数据增强的语音情感识别系统

Fig. 1 Speech emotion recognition system based on data enhancement

2 预处理

2.1 预处理——数据增强

数据增强对解决数据过拟合问题有极大的帮

助, 主要应用于现在的图像识别等领域, 而在语音识别领域的数据增强方法甚少。它是指通过增大训练的数据集来充分利用已有的数据集, 达到提高识别率的目的。目前图像识别领域常用的数

据增强方法有旋转、翻转变换、缩放变换、噪声扰动和尺度变换等。对原始数据库的图像进行改变来增加数据库样本。Cubuk 等^[9]提出了一种自动数据增强方法,使用搜索算法来寻找最佳策略,使得神经网络对目标数据集产生最高的验证精度,他们的方法在 cifar-10、cifar-100、svhn 和 imagenet 上达到了最高精度,其中在 imagenet 的数据集获得了 83.5% 的准确率。

研究语音样本的数据增强要考虑人类声音的特征。人类声音的频率范围主要分为男性和女性 2 部分。男声的声音分为男高音(Tenor)、男中音(Baritone)和男低音(Bass)。同样,女性的声音也可以分为女高音(Soprano)、女中音(Mezzo-soprano)和女低音(Alto)。其声音的实际频率范围如图 2 所示。研究表明,男低音、男中音与女低音、女中音的声调频率范围相同;女高音比男高音的频率略高;女性声音的平均频率大约是男性声音的平均频率的 1~1.75 倍。本文所采用的语音情感数据库为德国柏林 EMO-DB 数据库^[10],由 10 位不同性别(5 位男性、5 位女性)的专业人员录制。经过筛选得到了最终 535 条语句,共包含 7 种情感:生气(Anger)、悲伤(Sadness)、高兴(Happiness)、害怕(Fear)、中性(Neutral)、厌恶(Disgust)和无聊(Boredom)。根据男性声调与女性声调的不同之处,提出了一种基于时频域变换的方法进行数据增强。

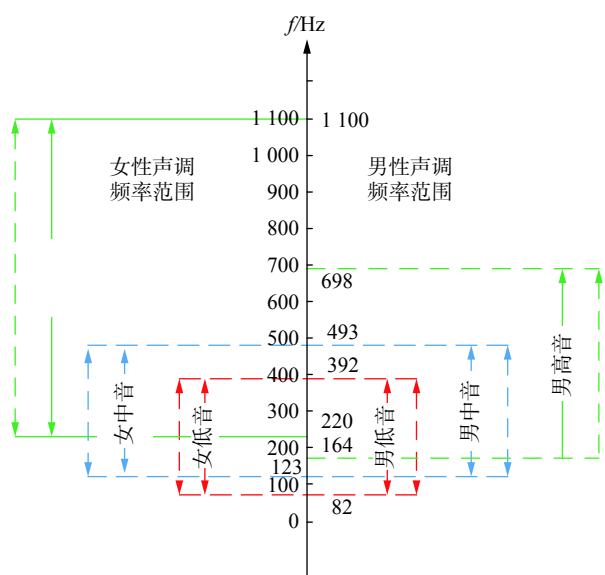


图 2 男性与女性声调频率范围

Fig. 2 Male and female tone frequency range

2.1.1 基于时域的数据增强方法

信号在时域上的 N 次抽取,也称为采样率压缩,会导致原始序列频谱的 N 倍扩展;信号在时域上的 P 次插值,也称为采样率扩张,会导致原

始序列频谱的 P 倍压缩。将抽取和插值结合起来,可以实现信号频率的 N/P 倍变换,当 $N > P$ 时,频率升高;当 $N < P$ 时,频率降低。时域信号的 N 次抽取原理如图 3 所示,设 $x(n)$ 为原始样本信号,经过 N 次抽取后生成 $x_n(n)$,那么 $x_n(n) = x(Nn)$ 。 $x(n)$ 的频谱为 $X(e^{j\omega})$, N 次抽取后频谱展宽为 $X_n(e^{j\omega})$ 。经过 N 次抽取变换后,频谱展宽。

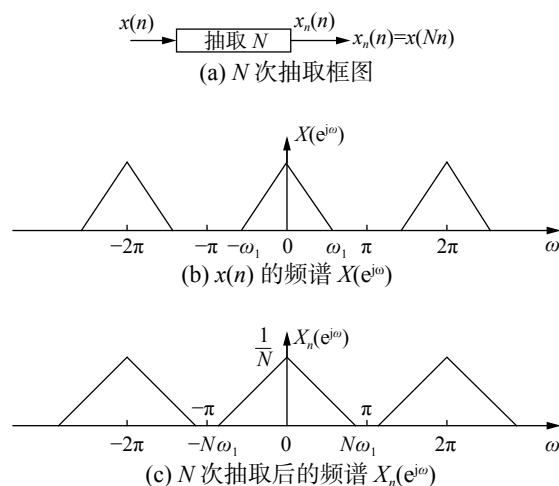


图 3 时域信号的 N 次抽取

Fig. 3 N -times extraction of time domain signals

同理,时域信号的 P 次插值如图 4 所示。 $x(n)$ 为原始样本信号,经过 P 次插值生成 $x_c(n)$ 。然而,进行抽取时,原始信号的频谱被周期拓展,如果原始信号最高频率大于抽取之后采样率的一半就会发生混叠(如果是复信号,则信号最高频率大于抽取之后采样率就会发生混叠)。 $h(n)$ 为防混叠滤波器。一般是先进行插值再进行抽取,以免造成频率混叠。 $x(n)$ 经过 P 次插值后生成信号 $x_p(n)$ 。其频谱 $X(e^{j\omega})$ 相对应的变换为 $X_p(e^{j\omega})$ 。根据图 5,抽取和插值可以实现信号频率的变换。选择合适的 P 和 N ,就能改变采样率。声音信号经过时域上的抽取和插值,改变了原始信号的频率。在此基础上,可以改变男性声调和女性声调的基本频率,达到男性声音和女性声音转换的目的。

2.1.2 基于频域的数据增强方法

频域变换的目的是改变原始声音样本的频谱,使男女声音互换。基音调整可以看作是频率的一种音阶变换。因为音调的高低是由频率决定的。频率越高,音调就越高。通过对信号进行快速傅里叶变换(fast fourier transformation, FFT),可以在频域内进行变换,将整个频谱向左或向右移动 100 Hz。然后进行反向快速傅里叶变换(inverse fast fourier transformation, IFFT)将音高返回

到时间域。男声和女声的基本频率可以通过左右变换来改变, 从而实现男声和女声的转换。

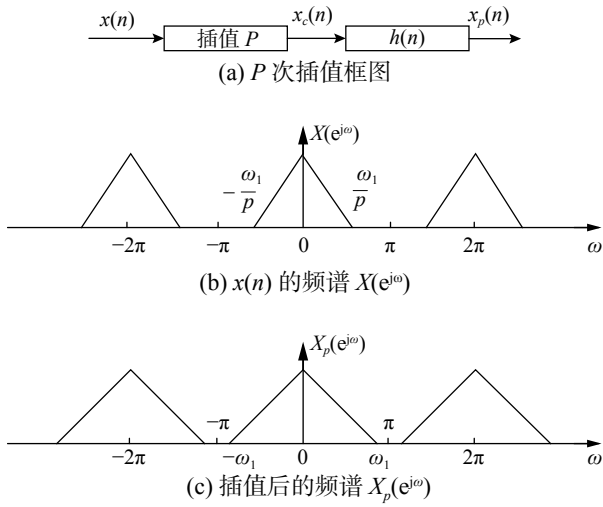


图 4 时域信号的 P 次插值

Fig. 4 P -times interpolation of time domain signals

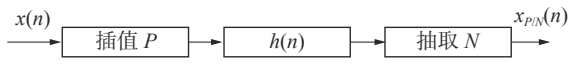


图 5 时域信号的 P/N 次变换

Fig. 5 P/N -times transformation of time domain signals

2.1.3 数据增强前后的语谱图比较

语谱图是一种以二维的平面表达三维信息的语音信号频谱图。其横轴代表时间, 纵轴代表频率, 颜色的深浅代表语音信号的强弱。图 6(a)、(b) 分别为未使用数据增强方法的样本语谱图和使用数据增强后的样本语谱图。使用了数据增强后的语谱图, 颜色清晰, 并未发现模糊或者样本损坏的情况。

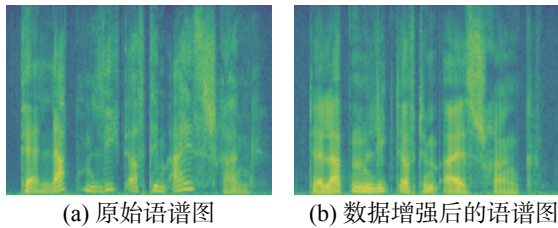


图 6 相同样本的语谱图比较

Fig. 6 Comparison of the spectrogram of same sample

2.2 特征提取

本文使用 openSMILE2.3.0 提取了 1 582 维的情感特征, 该特征集包含的 1 582 个特征是由 34 个低级描述特征和 34 个相应的变量作为 68 个低级描述特征轮廓值, 在此基础上应用 21 个函数得到 1 428 个特征, 另外, 对 4 个基于音高的低级描述特征及其 4 个变量系数应用了 19 个函数得到 152 个特征, 最后附加音高(伪音节)的数量和总体输入的持续时间(2 个特征)。此外, 还着重

分析了 10 个低级描述符的情感识别率, 主要包括: F0Final、F0finEnv、jitterDDp、jitterLocal、log-MelFreqBand、lsqFreq、pcm_fftMag_mfcc、pcm_loudness、shimmerLocal、voicingFinalUnclipped^[11]。

3 语音情感分类算法

在本节详细介绍使用的 3 种机器学习分类算法: 支持向量机、随机森林和 K 最邻近 (K -nearest neighbor)。

3.1 支持向量机

支持向量机是 Vapnik 等^[12-14] 在 20 世纪 90 年代提出的一种新的机器学习方法。它是一种广义的线性分类算法, 对数据按监督式学习进行二进制分类。支持向量机的基本思想是通过对非线性可分离样本进行非线性变换, 将其输入向量映射到另一个高维空间。为了在这个新空间中实现线性可分, 寻找最优分类超平面, 使超平面与不同类别的样本集之间的距离达到最大, 从而达到最大的泛化能力。它已经演变成 4 种类型:

1) 线性可分类型。

当想要分割 2 种数据类型时, 至少存在一个分割平面, 可以使 2 种类型完全分离, 这种类型没有边界最大化, 因此不能很好地控制错误, 称为线性可分类型。

2) 线性类型。

无法彻底地分割 2 种数据类型, 但可以使绝大多数的数据分割, 这种类型称为线性类型。

3) 非线性类型。

当一个类型不是线性可分的时候, 支持向量机通过提供一个“软边界”来处理数据分类, 即允许一组数据中的一些元素落在另一边, 但允许它们通过这个超平面而不产生任何重大异常, 允许系统变得不那么严格和更健壮。

4) 核函数。

核函数的目的是将原始被分类数据映射到更高维度的空间去进行分类, 将原本线性不可分类型在高纬度变成线性可分。

3.2 K 最邻近

在模式识别领域, K 最邻近算法是一种用于分类和回归的非参数统计方法。输入包含特征空间中 K 个最接近的训练样本。作为一种应用广泛的分类方法, K 最邻近拥有属性简单、无需参数、分类效果好的优点。

其分类的基本思想是: 在 K 最邻近分类中, 输出是一个分类组。一个物体的分类是由它的邻

居的“多数票”^[15-18]决定的。在 K 个最近邻中最常见的分类 (是一个正整数, 通常很小) 决定了分配给标签的分类。如果 $K=1$, 则对象的类别由最近的节点直接赋值。

3.3 随机森林

随机森林是一种基于“一组弱估计, 组合在一起, 形成一个强估计”原理的集成算法, 本文设计的随机森林分类器是一组 100 个决策树。让 X 和 Y 分别代表情绪的总数和样本的总数。为每个决策树选择一组引导程序样本。然后通过在每个节点上赋予 $y < Y$ 变量来构建决策树, 直到用尽所有的变量。在树的每个节点上, 用它们的变量拟合分类模型, 并找出截断点。经过训练, 未知样本的种类预测可以通过所有个体树的所有预测等级中的大多数投票来决定^[19-22]。

4 实验结果与分析

4.1 提取 1 582 维特征的识别结果

在这一部分, 主要介绍基于支持向量机、随机森林和 K 最邻近 3 种分类算法的识别率。对于实验数据, 本文使用柏林的情感语料库 EMO-DB。这是柏林理工大学录制的、德语情感语音数据库。采用 7 种情绪 (生气、无聊、厌恶、害怕、高兴、中性、悲伤) 模拟 10 种情绪 (男 5 例、女 5 例), 共计 800 句。经过筛选, 该语料库保留了 233 个男性情感句和 302 个女性情感句, 共计 535 个句子^[12]。本文利用数据增强来扩展原始柏林情感语料库, 将样本量增加到 700 个句子。以其中的 600 个句子作为训练集, 100 个句子作为测试集。

从图 7 得出, 基于支持向量机算法的平均识别率达到 86.65%。在这个模型中, 无聊、害怕和中性的识别率较高。然而, 厌恶的识别率很低。支持向量机算法是解决高维特征分类问题的有效方法。当特征维数大于样本数时, 仍然有较好的效果。核函数的数量很多, 可以很灵活地解决各种非线性分类回归问题。当样本量不是海量数据时, 分类精度高, 泛化能力强。

如图 8 所示, 基于随机森林的 1 582 维特征的识别率中, 中性和悲伤的识别率较高, 而厌恶的识别率很低, 只有 15.38%。随机森林的平均识别率为 61.14%, 与支持向量机算法相比较低。

如图 9 所示, K 最近邻算法的平均识别率为 79.38%。其中高兴的识别率很低, 只有 50%, 中性的识别率很高。因为 K 最邻近模型主要依赖于有限的相邻样本, 而不是依赖于判别类别域的方

法来确定类别。因此, 对于交叉或重叠较多的样本集合, K 最邻近方法比其他方法更适合。该算法更适用于样本量较大的类域的自动分类。然而, 小样本更容易出现分类错误。

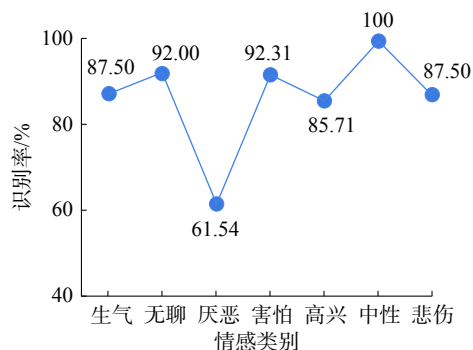


图 7 基于支持向量机的 1 582 维特征的识别率

Fig. 7 1 582 dimensional features recognition rate based on support vector machine

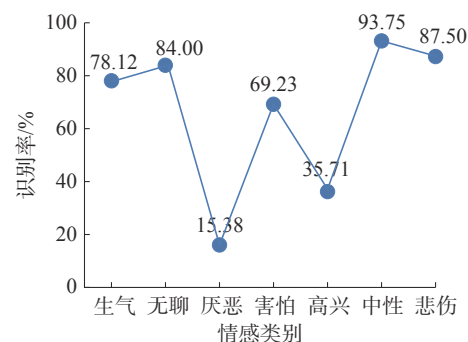


图 8 基于随机森林的 1 582 维特征的识别率

Fig. 8 1 582 dimensional features recognition rate based on random forest

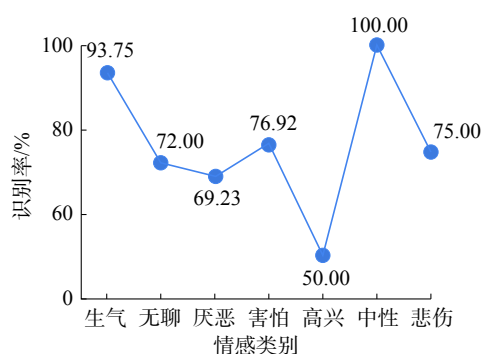


图 9 基于 K 最邻近的 1 582 维特征的识别率

Fig. 9 1 582 dimensional features recognition rate based on k nearest neighbor

图 10 为基于支持向量机、随机森林和 K 最邻近 3 种算法的 1 582 维特征的识别率比较。综上所述, 使用了数据增强方法的 3 种算法的中性情感的识别率普遍比较好; 而厌恶识别率相对较低。

表 1 为未使用数据增强和使用数据增强后的识别结果对比。表格纵轴代表参数为: 准确率、宏平均、微平均、召回率、 f_1 权重、汉明损失率、

kappa 系数和杰卡德相似系数等。对比发现,使用了数据增强后的支持向量机和 K 最邻近算法识别率提高了。其中,支持向量机算法的准确率提高了 2.53%, K 最近邻算法的准确率提高了 15.73%。然而,随机森林的识别结果较差,其汉明损失率增大了 13.1%。

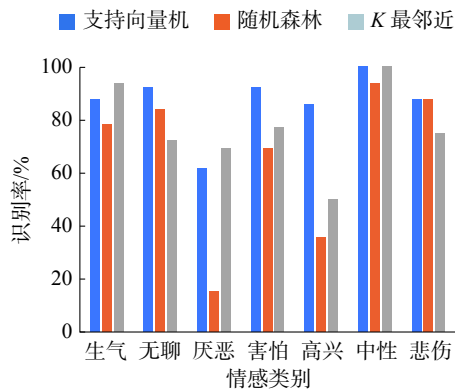


图 10 3 种算法的 1 582 维特征的平均识别率

Fig. 10 Average recognition rate of 1 582 dimensional features of three algorithms

表 1 未使用数据增强和使用后的识别结果对比

Table 1 Comparison of unused data enhancement and used

| 指标 | 支持向量机 | 随机森林 | K最邻近 |
|-------------|---------|---------|---------|
| 准确率 | 0.850 7 | 0.820 9 | 0.656 7 |
| | 0.876 0 | 0.689 9 | 0.814 0 |
| 宏平均 | 0.843 0 | 0.816 8 | 0.661 9 |
| | 0.864 0 | 0.652 4 | 0.806 3 |
| 微平均 | 0.850 7 | 0.820 9 | 0.656 7 |
| | 0.876 0 | 0.689 9 | 0.814 0 |
| 最小召回率 | 0.850 7 | 0.820 9 | 0.656 7 |
| | 0.876 0 | 0.689 9 | 0.814 0 |
| 最大召回率 | 0.846 8 | 0.797 6 | 0.623 8 |
| | 0.866 5 | 0.644 6 | 0.793 8 |
| f_1 权值 | 0.849 6 | 0.814 1 | 0.653 2 |
| | 0.874 4 | 0.671 2 | 0.808 2 |
| kappa系数 | 0.822 5 | 0.785 3 | 0.590 4 |
| | 0.852 4 | 0.627 0 | 0.777 2 |
| 汉明损失 | 0.149 3 | 0.179 1 | 0.343 3 |
| | 0.124 0 | 0.310 1 | 0.186 0 |
| J杰卡德相似系数归一化 | 0.850 7 | 0.820 9 | 0.656 7 |
| | 0.876 0 | 0.689 9 | 0.814 0 |
| 杰卡德相似系数值 | 114 | 110 | 88 |
| | 113 | 89 | 105 |

4.2 低级描述特征组的识别结果分析

表 2 为 10 组低级特征的识别结果分析,纵轴分别表示了基频 (F0final)、平滑的基频轮廓线 (F0finEnv)、帧间抖动 (jitter DDp)、局部抖动 (jitter-Local)、梅尔频带的对数功率 (LogMelFreqBand)、线谱对频率 (lspFreq)、梅尔频率倒谱系数 (pcm_fftMag_mfcc)、响度 (pcm_loudness)、震动 (shimmer) 和最终基频的可能发声概率 (voicingFinalUnclipped) 等。根据表 2 中数据统计, pcm_fftMag_mfcc 特征和 LogMelFreqBand 特征的识别结果较好。其中 pcm_fftMag_mfcc 特征在 3 种算法上的精确度分别为 74.63%、64.93% 和 66.42%; LogMelFreqBand 特征的精确度分别为 84.33%、73.13% 和 58.21%。面向数据增强方法的语音情感识别方法对 LogMelFreqBand 特征和 pcm_fftMag_mfcc 特征的识别性能有一定的提高,改善了传统语音情感识别方法识别率较低的状况。

表 2 10 组低级描述特征的识别比较

Table 2 Comparison of 10 sets of low-level descriptive features

| 低级描述特征 | 支持向量机 | 随机森林 | K最邻近 |
|-----------|---------|---------|---------|
| 基频 | 0.626 9 | 0.492 5 | 0.544 8 |
| 基频轮廓线 | 0.656 7 | 0.634 3 | 0.597 0 |
| 帧间抖动 | 0.395 5 | 0.335 8 | 0.343 3 |
| 局部抖动 | 0.403 0 | 0.335 8 | 0.365 7 |
| 梅尔频带对数功率 | 0.746 3 | 0.667 2 | 0.649 3 |
| 线谱对频率 | 0.694 0 | 0.567 2 | 0.656 7 |
| 梅尔频率倒谱系数 | 0.843 3 | 0.582 1 | 0.731 3 |
| 响度 | 0.507 5 | 0.507 5 | 0.485 1 |
| 震动 | 0.417 9 | 0.328 4 | 0.388 1 |
| 最终基频的发声概率 | 0.582 1 | 0.559 7 | 0.597 0 |

5 结束语

本文引入了一种基于时频域的数据增强方法,提出了将数据增强与语音情感识别结合的策略,并着重分析了与情感有关的特征和语音情感识别的关系。实验结果表明,使用了数据增强方法的支持向量机和 K 最邻近分类算法的识别率提高了,而随机森林的识别效果并不乐观;此外,改进了传统语音情感识别数据特征多且无关特征多的问题。

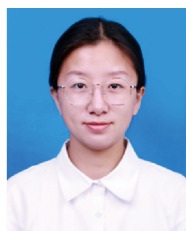
参考文献:

- [1] 吴雪, 宋晓茹, 高嵩, 等. 基于数据增强的卷积神经网络火灾识别 [J]. 科学技术与工程, 2020, 20(3): 1113–1117.
WU Xue, SONG Xiaoru, GAO Song, et al. Convolution neural network based on data enhancement for fire identification[J]. Science technology and engineering, 2020, 20(3): 1113–1117.
- [2] CHATZIAGAPI A, PARASKEVOPOULOS G, SGOUROPOULOS D, et al. Data augmentation using GANs for speech emotion recognition[C]//Proceedings of the 20th Annual Conference of the International Speech Communication Association. Graz, Austria, 2019.
- [3] ESCUDERO J P, NOVOA J, MAHU R, et al. An improved DNN-based spectral feature mapping that removes noise and reverberation for robust automatic speech recognition[J]. arXiv:1803.09016, 2018.
- [4] REN Fuji, MATSUMOTO K. Semi-automatic creation of youth slang corpus and its application to affective computing[J]. IEEE transactions on affective computing, 2016, 7(2): 176–189.
- [5] KAWASE T, NIWA K, HIOKA Y, et al. Automatic parameter switching of noise reduction for speech recognition[J]. Journal of signal processing, 2017, 21(2): 63–71.
- [6] YOUSEFI H, KANI A T, KANI I M, et al. Wavelet-based iterative data enhancement for implementation in purification of modal frequency for extremely noisy ambient vibration tests in Shiraz-Iran[J]. Frontiers of structural and civil engineering, 2020, 14(2): 446–472.
- [7] ELBAROUGY R, AKAGI M. Feature selection method for real-time speech emotion recognition[C]//Proceedings of the 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment. Seoul, Korea, 2017: 1–6.
- [8] MANGALAM K, GUHA T. Learning spontaneity to improve emotion recognition in speech[C]//Proceedings of the 19th Annual Conference of the International Speech Communication Association. Hyderabad, India, 2018.
- [9] CUBUK E D, ZOPH B, MANÉ D, et al. AutoAugment: learning augmentation strategies from data[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA, 2019.
- [10] CHEN Mingyi, HE Xuanji, YANG Jing, et al. 3-D convolutional recurrent neural networks with attention model for speech emotion recognition[J]. IEEE signal processing letters, 2018, 25(10): 1440–1444.
- [11] SCHULLER B W, STEIDL S, BATLINER A, et al. The INTERSPEECH 2010 paralinguistic challenge[C]//Proceedings of the 11th Annual Conference of the International Speech Communication Association. Makuhari, Chiba, Japan, 2010: 1342–6230.
- [12] BOSER B E, GUYON I M, VAPNIK V N. A training algorithm for optimal margin classifiers[C]//Proceedings of the 5th Annual Workshop on Computational Learning Theory. Pittsburgh, PA, USA, 1992: 144–152.
- [13] VAPNIK V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995.
- [14] VAPNIK V N. An overview of statistical learning theory[J]. IEEE transactions on neural networks, 1999, 10(5): 988–999.
- [15] 戴志诚, 李小年, 陈增照, 等. 基于 KNN 算法的可变权值室内指纹定位算法 [J]. 计算机工程, 2019, 45(6): 310–314.
DAI Zhicheng, LI Xiaonian, CHEN Zengzhao, et al. Variable-weight indoor fingerprinting localization algorithm based on KNN algorithm[J]. Computer engineering, 2019, 45(6): 310–314.
- [16] LARIJANI M R, ASLI-ARDEH E A, KOZEGAR E, et al. Evaluation of image processing technique in identifying rice blast disease in field conditions based on KNN algorithm improvement by K-means[J]. Food science & nutrition, 2019, 7(12): 3922–3930.
- [17] SATHISHKUMAR R, KALAIARASAN K, PRABHAKARAN A, et al. Detection of lung cancer using SVM classifier and KNN algorithm[C]//Proceedings of 2019 IEEE International Conference on System, Computation, Automation and Networking. Pondicherry, India, 2019.
- [18] 连天友, 余勤. 改进 KNN 算法对人体身份的识别 [J]. 计算机工程与应用, 2019, 55(11): 142–146, 243.
LIAN Tianyou, YU Qin. Human identity recognition using improved KNN method[J]. Computer engineering and applications, 2019, 55(11): 142–146, 243.
- [19] PAUL A, MUKHERJEE D P, DAS P, et al. Improved random forest for classification[J]. IEEE transactions on image processing, 2018, 27(8): 4012–4024.
- [20] YEŞILKANAT C M. Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm[J]. Chaos, solitons & fractals, 2020, 140: 1–8.
- [21] BUTT A M, BHATTI Y K, HUSSAIN F. Emotional speech recognition using SMILE features and random

forest tree[M]. BI Yaxin, BHATIA R, KAPOOR S. Intelligent Systems and Applications. Cham: Springer, 2020.

- [22] DAI Jingzhao, ZHANG Yaan, HOU Jintao, et al. Sparse wavelet decomposition and filter banks with CNN deep learning for speech recognition[C]//Proceedings of 2019 IEEE International Conference on Electro Information Technology. Brookings, SD, USA, 2019.

作者简介:



李茜茜, 硕士研究生, 主要研究方向为语音情感识别和特征处理。



沈晓燕, 教授, 博士, 南通市“226”工程二级中青年科技领军人才、南通市康复医学会康复教育专业委员会委员、南通大学信息科学与技术学院信息与通信工程专业医学信息技术学科带头人, 主要研究方向为生物神经接口技术、神经信号检测电路和功能电激励电路设计、神经信号和肌电信号采集技术与分析、神经信号再生和功能重建。发表学术论文 40 余篇。



任福继, 教授, 博士, 日本工程院院士和欧盟科学院院士, 中国人工智能学会名誉副理事长, 日本工学会、IE-ICE、CAAI Fellow, 日本国际先进信息研究所主席, 获吴文俊人工智能科学技术奖创新一等奖等, 主要研究方向为人工智能、情感计算、自然言语理解、模式识别。申请发明专利 10 余项。发表学术论文 500 余篇。

第四届模式识别与计算机视觉大会

The 4th Chinese Conference on Pattern Recognition and Computer Vision

第四届中国模式识别与计算机视觉大会将于 2021 年 10 月 29 日至 11 月 1 日在北京举行。会议由中国图象图形学学会(CSIG)、中国人工智能学会(CAAI)、中国计算机学会(CCF)和中国自动化学会(CAA)联合主办; 由北京科技大学、北京交通大学和北京邮电大学共同承办, 中山大学和清华大学协办, 是国内顶级的模式识别和计算机视觉领域学术盛会。本届会议将主要汇聚国内国外模式识别和计算机视觉理论与应用的广大科研工作者及工业界同行, 共同分享我国模式识别与计算机视觉领域的最新理论和技术成果, 提供精彩的学术盛宴。现向广大科技工作者公开征集高质量、原创性的优秀英文学术论文。大会录用的稿件将在会上展示, 会议论文集将由 Springer 出版社出版, 并被 EI 和 ISTP 检索; 优秀的论文将推荐到国内外高质量期刊发表。

1. 征文范围(包括但不限于)

模式分类与聚类分析、结构模式识别、机器学习、神经网络与深度学习、特征提取与特征选择、计算机视觉基础理论、底层视觉理解、图像处理、三维视觉与重构、文档分析与识别、字符识别、人脸识别与姿态识别、目标检测、跟踪与识别、行为识别、多媒体分析与推理、医学图像处理与分析、生物特征识别、遥感影像解译优化及学习方法、多模态信息处理、性能评测和基准数据库、视频分析与理解、视觉应用与系统、机器人、自动驾驶中的视觉问题。

2. 重要日期

投稿截止日期: 2021 年 4 月 30 日

录用通知日期: 2021 年 6 月 30 日

终稿提交日期: 2021 年 8 月 15 日

会议举办日期: 2021 年 10 月 29 日至 2021 年 11 月 1 日

3. 会议投稿

<https://cmt3.research.microsoft.com/PRCV2021/>