

语音情感识别中面向小数据集的 CGRU 方法

吴俊清,倪建成,魏媛媛

(曲阜师范大学 软件学院,山东 曲阜 273100)

摘要:为增强人机交互的和谐,提升语音情感识别的精度,提出一种面向小数据集的 CGRU 深度学习方法。该方法将原始音频通过上移和下移操作进行语音增强,将增强后的语音信号映射到 Mel 尺度并生成 Mel 功率谱图,然后对其做旋转、切角、偏移等图像增强操作,并结合卷积神经网络(CNN)对频域特征的捕捉能力和门控循环单元(GRU)网络对时序信息的特性获取能力构成融合模型 CGRU,该模型通过自动学习深度谱特征进行情感识别。实验分别验证了利用谱图特征与手工特征在 Emo-DB 上的识别效果,并比较了 CLSTM 与 CGRU 的时间性能。结果表明,利用谱图特征在 CGRU 方法上的情感识别精度达到 98.39%,超过传统手工特征 eGeMAPS 在该数据库上的识别效果,提出的方法在语音情感识别任务上获得有竞争力的表现。另外,在相同的训练参数下,CGRU 要比 CLSTM 具有更加良好的时间性能。

关键词:语音情感识别;卷积神经网络;门控循环单元;小数据集;谱图特征

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2020)12-0077-06

doi:10.3969/j.issn.1673-629X.2020.12.014

CGRU Method for Small Datasets in Speech Emotion Recognition

WU Jun-qing, NI Jian-cheng, WEI Yuan-yuan

(School of Software, Qufu Normal University, Qufu 273100, China)

Abstract: In order to enhance the harmony of human-computer interaction and improve the accuracy of speech emotion recognition, a CGRU deep learning method for small datasets is proposed. In this method, the original audio is enhanced by moving up and down. The enhanced speech signal is mapped to the Mel scale to generate a Mel power spectrum, and then image enhancement operations such as rotation, chamfering and shifting are performed. Combining the ability of the convolutional neural network (CNN) to capture the frequency domain features and the ability of the gated recurrent unit (GRU) network to acquire the time sequence information, a fusion model CGRU is formed. This model automatically learns deep spectrum features to perform emotion recognition. The experiments verify the recognition effect of spectral features and manual features on Emo-DB, and compare the time performance of CLSTM and CGRU. The results show that the accuracy of emotion recognition using the CGRU method of spectral features reaches 98.39%, which exceeds the recognition effect of the traditional manual feature eGeMAPS. The proposed method achieves competitive performance on speech emotion recognition tasks. In addition, under the same training parameters, CGRU has better performance of time than CLSTM.

Key words: speech emotion recognition (SER); convolutional neural networks; gated recurrent unit; small datasets; spectral feature

0 引言

情感是提升人机交互和谐的重要因素,也是影响人类行为决策的内在驱动。语音中不仅包含大量的语义信息,还蕴含着丰富的情感内涵。语音情感识别的优劣很大程度依赖于特征的构建,而情感作为一种抽象的个体,很难有标准的特征表示,因此如何获取准确的情感特征成为语音情感识别的一大难题。传统的语音特征大致从韵律学特征(如基频、能量、时长)、频谱特征(Mel 倒谱系数-MFCC、线性预测倒谱系数-

LPCC)和声音质量特征(即人们赋予语音的主观评价指标,用于衡量语音是否纯净、清晰、容易辨识等,如喘息、颤音、哽咽等)^[1]提取关联信息,这些特征都是手工构建,需花费大量时间和精力,因此,人们逐渐将注意力放在自动提取特征方面。

谱图可以捕捉语音信号间不同的情感特征,是一种流行的特征表示方式,研究表明,使用深度学习方法从谱图图像中自动学习情感特征可以达到甚至超过手动特征的识别效果。文献[2-3]都构建了一种包含静

收稿日期:2019-12-28

修回日期:2020-04-29

基金项目:国家自然科学基金青年项目(61601261);山东省研究生教育质量提升计划项目(SDYY17136)

作者简介:吴俊清(1994-),女,硕士,研究方向为语音情感识别、机器学习;倪建成,博士,教授,CCF 高级会员(16502S),研究方向为分布式计算、机器学习、数据挖掘。

态系数、一阶系数和二阶系数的三通道的 log-mel 谱图,文献[2]中使用谱图提取的深度特征较使用韵律和 MFCC 特征的平均召回率提升了 1.7%。文献[4]中使用 mel 谱图提取深度谱特征,发现利用 BLSTM-FCN 网络模型计算出的每个频带的深谱特征与传统的语音情感识别的声学特征表示方法相比,在性能上是相当的。

1 相关工作

近年来,卷积神经网络在图像识别任务中表现优异^[5-7],因此,利用 CNN 捕捉局部特征的能力实现自动获取识别任务特征,从而代替传统手工特征构建,简化特征提取过程的思想逐渐流行。CNN 具有局部连接、权值共享的特性,使模型对平移缩放具有一定的不变性,在各种信号和信息处理任务中的性能优于标准的全连接神经网络^[8]。由于它们高精度的识别率,因此在大规模图像识别任务中成为最主要的方法。Mao 等人^[9]首次使用卷积神经网络(CNN)学习 SER 的情感显著性特征,并在几个基准数据集上显示出优异的性能。文献[10]通过预训练的 AlexNet 作为特征提取器对信号生成的谱图提取深度特征,以类似的方式,文献[11]依据 caffeNet 为基准结构删减掉原始三个连续卷积层中的两个,并以灰度谱图作为网络输入,减小训练时间加速模型收敛。

CNN 可以有效地获取图像高级特征,但受限于卷积网络自身结构特性,因此很难捕捉到时域方向的特征,近期文献表明,研究者更倾向于构建一种结合 CNN 和 RNN 的新型网络结构,以此更好地学习谱图时域和频域特征。在文献[12-13]中,研究人员将卷积层和递归层结合起来,以提高基于时域语音信号或原始信号的语音情感识别任务的性能。文献[14]提出了一种结合 CNN 和 LSTM 且用于大量词汇语音识别的卷积递归神经网络(Crnn)。实验结果表明,该模型具有较高的语音识别性能。与此同时,Aharon Satt 等人^[15]构建了一种三层卷积连接一层 LSTM 的深层网络结构用于语音情感识别,该系统与以前没有延迟约束(约束 ≤ 3 秒)的工作相比,在通用基准测试数据集 IEMOCAP 上达到最先进的识别精度。文献[16]提出一种基于串联 CNN 和 RNN 训练的 SER 方法,该方法无需使用任何传统的手工功能。将所提方法应用于语音情感数据库中,分类结果证明比使用常规分类方法具有更好的准确性。

虽然利用深度学习提取谱图特征成为语音情感识别的一种趋势,但是应用于小数据集上的深度学习研究相对较少,为解决小数据集上存在深度学习过拟合且识别精度低的问题,提出一种基于语音和图像双重

增强的卷积神经网络与门控循环单元网络结合的深度学习模型。通过对 Emo-DB 德语情感语料库进行语音和图像增强,对增强后的 Mel 功率谱图进行深度学习的模型构建并对识别结果进行评估。

2 模型

2.1 模型结构

2.1.1 卷积模块

近年来,卷积神经网络在图像识别领域大放异彩,得益于卷积操作强大的特征提取能力,将增强后的语音生成的 Mel 功率谱图作为卷积神经网络的输入。不同于传统卷积神经网络的结构,在构建前两层卷积层时用 BN 层代替池化层,这样可以尽可能多地保留局部信息特征,减少池化操作带来的信息丢失,加入的 BN 层可以加速模型训练并提高泛化性能。尽管卷积神经网络会学习到不错的抽象特征,但会忽略时间方向的信息,考虑到这个问题,在卷积模块后面接入 GRU 模块,以便获取时域信息。

2.1.2 GRU 模块

RNN 在反向传播过程中不断进行链式求导,后期会出现梯度消失的现象,导致网络很难学习到距离较远的单元内容。为了解决这一问题,LSTM 和 GRU^[17](门控循环单元)相继而生。GRU 是 LSTM 的一种变体,较 LSTM 相比只有两个“门”,单元结构如图 1 所示。重置门 r_t 决定 $t-1$ 时刻信息遗忘的程度,计算方式如式(2)所示, r_t 的值越小说明前一时刻的信息被忽略得越多,其中 σ 表示“sigmoid”函数;更新门 z_t 表示 $t-1$ 时刻的信息保留到当前时刻的程度,当 z_t 越大时,说明前一时刻信息被保留得越多,以此解决递归神经网络中存在的长依赖问题。最终,由式(4)可知,当前时刻信息 h_t 由前一时刻信息 h_{t-1} 及当前时刻候选集 \tilde{h}_t 决定。GRU 较 LSTM 模型结构更简单,参数更少,因此更容易训练。

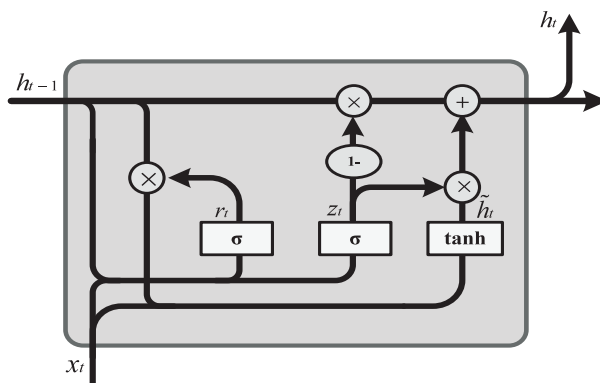


图 1 GRU 单元结构

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (1)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (2)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad (3)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (4)$$

2.1.3 CGRU 模型

CGRU 模型由卷积模块与 GRU 模块构成,该模型结构如图 2 所示。卷积模块利用卷积操作提取 Mel 功率谱图的抽象特征,然后经过 BN 层对特征进行规范化操作,使该层数据尽量接近标准分布,再通过 ReLU 层将数据映射到 $[0, \infty)$,由于该激活函数不存在“饱和区”,因此可有效避免梯度消失。该结构只在最后一层卷积后加入池化层,可适当地减少特征维度,提升模型训练速度。

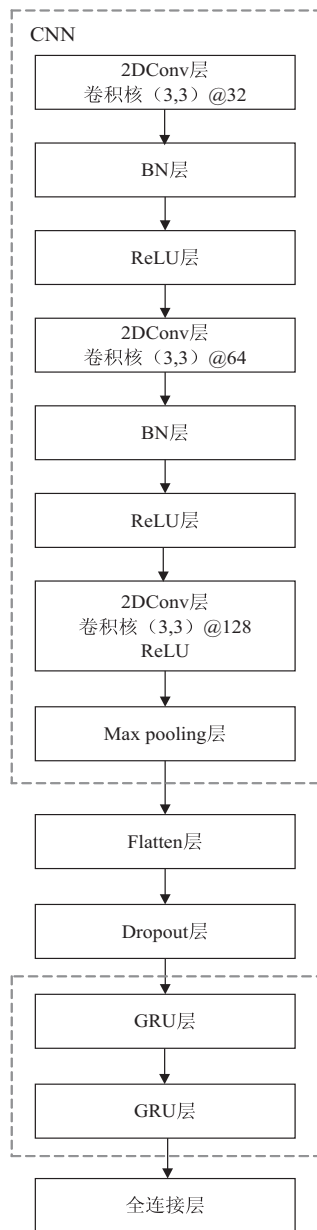


图 2 CGRU 结构

卷积模块后接 Flatten 层将多维输入转换成一维,再连接 dropout 层,训练时使部分神经元失活,可有效避免过拟合发生。最后,通过两层 GRU 获取时间序列

层面所包含的信息,对特征作进一步提取,提取后的特征作为全连接层的输入,经过 softmax 计算出类别概率。

2.2 语音预处理

Emo-DB 共有 535 个音频文件,数据存在数据量小且类别不平衡的现象,因此并不适合直接作为深度学习的学习对象。考虑到每个音频时长均不超过三秒,因此,在保持时长不变的前提下对音频进行上移和下移操作以此增强语音数据的丰富性。数据原始采样率为 16 kHz,论文重采样至 22.05 Hz。音频增强具体操作步骤如下(文中涉及到的上移/下移操作表示分别做上移和下移操作的简称):

(1)对“angry”类别的所有音频分别做上移/下移 1 个半步、上移/下移 2 个半步、上移/下移 2.5 个半步操作;

(2)对其余六类音频均做上移/下移 1 个半步、上移/下移 1.5 个半步、上移/下移 2 个半步、上移/下移 2.5 个半步、上移/下移 3.5 个半步操作;增强后的音频数量如表 1 所示。

表 1 EMO-DB 增强前后音频数量

情感类别	Angry	boredom	disgust	fear	happy	neutral	sad	总计
原始音频(个)	127	81	46	69	71	79	62	535
增强音频(个)	889	891	506	759	781	869	682	5 377

2.3 Mel 功率谱图生成

Mel 功率谱图是原始信号映射到时间和频率区域的一种二维表示方式,文中使用 Python 的音频分析工具包—librosa 实现绘制。首先对原始音频信号 y_i 进行分帧操作并使用汉宁窗进行加窗;然后按照式(5)逐帧进行短时傅里叶变换(STFT),STFT 通过在短重叠窗口上计算离散傅里叶变换(DFT)来表示时频域中的信号。函数中 FFT 窗口大小为 2 048 个样本点,对应 22 050 Hz 采样率下 93 ms 的持续时间;重叠样本 512 个,相当于 22 050 Hz 采样率下 23 ms 的持续时间;对得到的短时傅里叶变换矩阵 S (是包含帧 t 处频率 f 大小的复值矩阵)按照式(6)计算功率谱,其中 M 为 $[m, k/2+1]$ 大小的 Mel 变换矩阵, m 代表 Mel 过滤器的个数, k 表示 fft 组件的个数,将信号进行 Mel 变换更贴合人耳对声音响度的听觉特性;最终依照式(7)将功率谱图转化为分贝(db)单位。

$$S[f, t] = \text{STFT}\{y_i\} \quad (5)$$

$$P = M * |S|^2 \quad (6)$$

$$P_{db} = 10 \log_{10}(P/\max(S)) \quad (7)$$

3 实验设计

该实验在 Ubuntu 16.04.5 LTS 环境下运行,使用 Python 版本为 3.6.2,在两块 NVIDIA Tesla P100 的 GPU 上训练三天半。实验使用 keras 框架实现,且后端为 tensorflow。

3.1 数据库

Emo-DB(柏林语料库)是德文情感语料库,由柏林工业大学的 10 位专业演员录制而成。该数据库包含 7 类常见情感(愤怒、无聊、沮丧、恐惧、愉悦、中性、悲伤)的 535 个音频文件,且全部为 .wav 无损格式,采样率由最初的 44.1 kHz 下采样至最终的 16 kHz。文中实验时将语料库均采样为 22.05 kHz。

3.2 图像预处理

通过增强后的音频生成的 Mel 谱图共 5 377 张,较原始的数量扩展了 9 倍,为进一步增强数据的丰富性,避免过拟合现象的发生,现对 7 类数据中的每一张谱图做增强处理。首先,对图像进行归一化操作,使像素值限定在 0~1 之间,避免数值过大在传播的过程中发生梯度消失;再次,对每一张图片做逆时针剪切 0.2 度角、图像放大 0.2、图像垂直偏移 0.1、图像水平偏移 0.1、通道偏移 10 的操作,至此,每张图片增强为原来的 5 倍。该操作通过 keras 框架的 ImageDataGenerator 类完成。

3.3 eGeMAPS 特征集

为对比深度谱图特征与手工声学特征在语音情感识别方面的表现,提供 eGeMAPS 特征集用于语音情感识别任务。eGeMAPS 作为语音情感识别中经典的特征集,是 GeMAPS 的扩展特征集,该表示特征通过 openSMILE^[18] 开源工具包提取,包含频率、能量、光谱、倒谱和动态信息在内的 88 个高级统计特征。

3.4 实验设置

实验将 5 337 张 Mel 谱图按 7:1.5:1.5 的比例划分为训练集、验证集和测试集,划分后的数量为 3 761、810 和 806。该文对输入格式、卷积核数量、训练批次大小等逐一进行设置,模型的超参数细节如下:

(1)输入的 Mel 功率谱图为 128 * 128 大小的三通道彩色图片,且全部为 .png 格式。

(2)根据卷积层特征提取特点,浅层卷积层关注局部的信息特征,深层卷积层关注更加细致的整体特征,因此滤波器个数分别以 32、64 和 256 的顺序递增设置,以适应网络特性更好地提取特征。

(3)设置 3 * 3 的卷积核,步长为 1,最大池的大小设为 2 * 2。

(4)GRU 共两层,单元个数依次为 126 和 35。

(5)Dropout 系数设为 0.38,使用 Adam 优化器进行优化,初始学习率为 0.000 1,该值过大会导致训练

过程中发生震荡现象, $\beta_1=0.9$, $\beta_2=0.999$ 。batch_size 设为 92,迭代次数为 300 次。

4 实验分析及评估

实验将 Mel 功率谱图作为 CGRU 的输入,利用 CNN 学习谱图的频域信息并结合 GRU 对序列信息的捕捉特性学习谱图的时域特征,从而加强对 Mel 功率谱图的特征提取效果。对 EMO-DB 在 CGRU 上的表现情况如图 3 所示。该数据集在提出的方法上表现出了不错的性能,训练集通过迭代训练使精度收敛到 0.9 以上,损失值稳定至 0.1 左右。使用验证集对训练模型的效果进行验证,发现验证集的精度和损失值基本与训练集的收敛趋势保持一致,精度由 0.301 2 上升至 0.965 4,损失值由 1.781 1 减小至 0.095 5,该结果显示出了模型较好的拟合能力。最后,对该模型在测试集上的精度和损失值进行计算作为最终的评估结果,实验发现该模型在测试集上也具有良好的性能,结果显示 CGRU 在 EMO-DB 上的精度达到 98.39%,损失减小到 0.048。

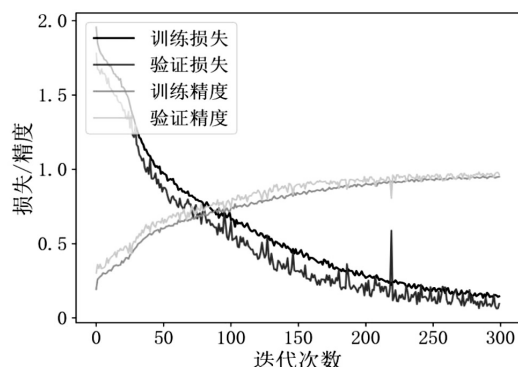


图 3 EMO-DB 损失/精度

文中在相同参数设置下对 CLSTM 与 CGRU 的时间性能做了对比,考虑到时序模型存在梯度消失问题,因此,将 LSTM 和 GRU 层设为两层。如表 2 所示,在相同的模型参数下,CGRU 用时 3.64 天完成训练,而 CLSTM 用了 6.22 天,比前者多出近一倍时间,这也符合 LSTM 与 GRU 模型自身的特性,由于 GRU 比 LSTM 少了一个门操作,因此在处理大量数据时时间要优于 LSTM。

表 2 CLSTM 与 CGRU 时间对比

方法	结构	时间/天
CLSTM	CNN+2 层 LSTM	6.22
CGRU	CNN+2 层 GRU	3.64

为了验证传统手工特征与深度谱特征在语音情感识别中的作用,在两组特征集上做了四组对比试验,实验结果如表 3 所示。可以看出,在使用完全相同的训练集与测试集的情况下,谱图特征展现出了有竞争力

的表现,两种方法下的谱图特征均达到了 90% 以上的识别率。XGBoost 作为一种强大的集成算法,相比随机森林而言,对数据不太敏感,模型更具鲁棒性,在 eEeMAPS 特征集上比随机森林表现出更优的性能,精度高出 11.79%。实验显示,利用深度学习方法获取到的谱图特征在一定程度上接近甚至超越传统手工特征的识别精度。

表 3 四种方法在两种特征下的识别精度

序号	方法	特征	精度/%
1	随机森林	eGeMAPS 手工特征	82.75
2	XGBoost	eGeMAPS 手工特征	94.54
3	CLSTM	Mel 功率谱图特征	91.94
4	CGRU	Mel 功率谱图特征	98.39

CGRU 利用 Mel 功率谱图在深度学习方法中对 EMO-DB 情感的识别精度达到 98.39%,超过了文献[19]中 95.1% 的准确率,也比方法 2 的精度高出了 3.85%。虽然使用 Mel 功率谱图特征的 CLSTM 识别率低于使用手工特征的 XGBoost,但为了便于比较

CLSTM 与 CGRU 的模型表现,统一将迭代次数设置为 300,由于 LSTM 较 GRU 收敛速度较慢,因此方法 3 还可继续收敛,说明了深度学习下的谱图特征较传统手工特征具有不错的表征能力。

实验使用混淆矩阵对柏林语料库的 Mel 功率谱图特征在 CGRU 模型上的效果进行评估,矩阵值为归一化后结果,其中,横向代表模型的“预测标签”,纵向代表数据的“真实标签”,对角线数值表示正确预测类别的数量占预测总类别数量的比率,深色部分表示预测效果良好,浅色部分表示预测结果欠佳,结合混淆矩阵的特性,实验结果应在对角线位置呈最深颜色以达到最佳效果。

由图 4 可知,“neutral”的预测效果最好,而“disgust”的预测效果最差,且大部分被误分为“fear”类,其主要原因为该两种情感在情绪的“动机-维度”模型中均属于高动机低效价的情感,二者在情感表达方面具有一定的相似度,因此在模型识别时会有一定程度的混淆。



图 4 EMO-DB 归一化混淆矩阵

5 结束语

借鉴深度学习的思想,使用深度谱特征代替手工特征,提出在小数据集上完成语音情感的深度学习建模。利用卷积神经网络强大的表征能力,将生成的 Mel 功率谱图作为 CNN 的输入获取语音的频域特征,再结合语音信号的时序特性,加入 GRU 模型获取谱图的时域特征,以此构建一种 CGRU 的融合模型对小数据集进行深度学习。实验在 EMO-DB 数据集上利用谱图特征与传统语音情感识别的声学手工特征的识别结果相比具有良好的竞争力。该方法摒弃了传统构建特征的繁琐过程,使用深度学习自动提取特征。

未来,将注意力放在融合特征方面,结合深谱特征的潜在信息和声学特征的专业特性构造鲁棒性更加优越的融合特征,以适应更广泛的数据集。

参考文献:

- [1] 韩文静,李海峰,阮华斌,等. 语音情感识别研究进展综述[J]. 软件学报,2014,25(1):37-50.
- [2] ZHANG Shiqing, ZHANG Shiliang, HUANG Tiejun, et al. Speech emotion recognition using deep convolutional neural network and discriminant temporal pyramid matching[J]. IEEE Transactions on Multimedia, 2018, 20(6):1576-1590.
- [3] 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.
- [4] ZHAO Ziping, ZHAO Yiqin, BAO Zhongtian, et al. Deep spectrum feature representations for speech emotion recognition[C]//ASMM- MMAC' 18; 2018 ACM MM workshop. New York: ACM, 2018:27-33.

- [5] 刘万军,梁雪剑,曲海成. 基于双重优化的卷积神经网络图像识别算法[J]. 模式识别与人工智能,2016,29(9):856-864.
- [6] 唐贤伦,杜一铭,刘雨微,等. 基于条件深度卷积生成对抗网络的图像识别方法[J]. 自动化学报,2018,44(5):855-864.
- [7] 郑远攀,李广阳,李 晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用,2019,55(12):20-36.
- [8] 周飞燕,金林鹏,董 军. 卷积神经网络研究综述[J]. 计算机学报,2017,40(6):1229-1251.
- [9] MAO Qirong, DONG Ming, HUANG Zhengwei, et al. Learning salient features for speech emotion recognition using convolutional neural networks[J]. IEEE Transactions on Multimedia,2014,16(8):2203-2213.
- [10] CUMMINS N, AMIRIPARIAN S, HAGERER G, et al. An image-based deep spectrum feature representation for the recognition of emotional speech [C]//Proceedings of the 2017 ACM on multimedia conference. [s. l.]: ACM,2017:478-484.
- [11] PPAKOSTAS M, GIANNAKOPOULOS T. Speech-music discrimination using deep visual feature extractors[J]. Expert Systems with Applications,2017,114:334-344.
- [12] KEREN G, SCHULLER B. Convolutional RNN: an enhanced model for extracting features from sequential data [C]//2016 international joint conference on neural networks. [s. l.]: IEEE,2016:3412-3419.
- [13] TRIGEORGIS G, RINGEVAL F, BRUECKNER R, et al. A-dieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network [C]//2016 IEEE international conference on acoustics, speech and signal processing. Shanghai: IEEE,2016:5200-5204.
- [14] LI Xiangang, WU Xihong. Long short-term memory based convolutional recurrent neural networks for large vocabulary speech recognition [C]//16th annual conference of the international speech communication association. Dresden, Germany: Technische Universität Berlin,2015:3219-3223.
- [15] SATT A, ROZENBERG S, HOORY R. Efficient emotion recognition from speech using deep learning on spectrograms [C]//INTERSPEECH 2017:18th annual conference of the international speech communication association. Stockholm, Sweden: Stockholm University,2017:1089-1093.
- [16] LIM W, JANG D, LEE T. Speech emotion recognition using convolutional and recurrent neural networks [C]//APSIPA: 2016 Asia-Pacific signal and information processing association annual summit and conference. Jeju, South Korea: IEEE,2016.
- [17] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]//Proceedings of the 2014 conference on empirical methods in natural language processing. Stroudsburg, PA: ACL,2014:1724-1734.
- [18] EYBEN F, WENINGER F, GROSS F, et al. Recent developments in opensmile, the munich open-source multimedia feature extractor [C]//Proceedings of the 21st ACM international conference on multimedia. [s. l.]: ACM,2013:835-838.
- [19] PAN Y, SHEN P, SHEN L. Speech emotion recognition using support vector machine [J]. International Journal of Smart Home,2012,6(2):101-108.
- +++++
- (上接第 76 页)
- [13] LUO W, TAY W P, LUO W, et al. Estimating infection sources in a network with incomplete observations [C]//Global conference on signal & information processing. Austin, TX, USA: IEEE,2013:301-304.
- [14] ZANG Wenyu, ZHANG Peng, ZHOU Chuan, et al. Locating multiple sources in social networks under the SIR model: a divide-and-conquer approach [J]. Journal of Computational Science,2015,10:278-287.
- [15] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks [J]. Proceedings of the National Academy of Sciences of the United States of America,2002,99(12):7821-7826.
- [16] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks [J]. Physical Review E,2004,69:026113.
- [17] CLAUSET A, NEWMAN M E J, MOORE C. Finding community structure in very large networks [J]. Physical Review E,2004,70(6):066111.
- [18] MASSEN C P, DOYE J P K. Identifying communities within energy landscapes [J]. Physical Review E,2005,71(4):046101.
- [19] NEWMAN M E J. Finding community structure in networks using the eigenvectors of matrices [J]. Physical Review E,2006,74(3):036104.
- [20] 崔宇童. 基于谱聚类优化的社区发现方法研究 [D]. 徐州: 中国矿业大学,2018.
- [21] 唐风琴,丁文文. 基于谱聚类带有节点特征的社区发现算法 [J]. 中国科学技术大学学报,2018,48(2):161-172.
- [22] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks [J]. Science,1999,286(5439):509-512.
- [23] WATTS D J, STROGATZ S H. Collective dynamics of small world networks [J]. Nature,1998,393(6684):440-442.