语音情感识别的应用和挑战

SSIL_SZT_ZS 发表于 2021/08/26 17:02:18

【摘要】 本文介绍了语音情感识别领域的发展现状，挑战。重点介绍了处理标注数据缺乏的问题，并讲解了我们自己的语音情感识别方案。

语音情感识别的应用和挑战

情感在人与人的交流中扮演者重要的角色。情感识别具有极大的应用价值，成功的检测人的情感状态对于社交机器人、医疗、教育质量评估和一些其他的人机交互系统都有着重要意义。本文的要点有：
1、情感识别的基础知识和应用场景。
2、语音情感识别技术的介绍以及面临的挑战。
3、如何解决数据缺乏问题，我们的方案是什么。

1. 什么是情感识别?

情感是人对外部事件或对话活动的态度。人的情感一般分为：高兴、生气、悲伤、恐惧和惊喜等。机器对采集的信号进行分析，从而得到人的情感状态，这一过程就是情感识别。通常，能用来进行情绪识别的信号包括两个方面，一个是生理信号如呼吸、心率和体温，另一个是行为表现包括面部表情、语音和姿态等等。人脸与语音得益于简单的采集方式，经常被用来识别对象的情感。情感识别能帮助系统了解对象的情感状态以及其对某个话题或事务的态

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发

【模型评估(一)】AI市场模型评估端到端流程打通

基于sentinel 1 和 2 数据的逐月Kmeans聚类分类的NDVI逐月统计

GEE ——多种机器学习方法（随机森林、cart、svm等）进行土地分类用光...

垃圾分类之场景化AI体验【玩转华为云】

ModelArts域适应算法EfficientMixGVB

对cann的算子编译注册工程的说明解析【对应工程1_custom_op】

AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...

2022年年终总结：日出万物生，日落满天星

GEE：社会联系指数（SCI）Social Connectedness Index数据集

垃圾分类之场景化AI体验【玩转华为云】

ModelArts域适应算法EfficientMixGVB

对cann的算子编译注册工程的说明解析【对应工程1_custom_op】

AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...

2022年年终总结：日出万物生，日落满天星

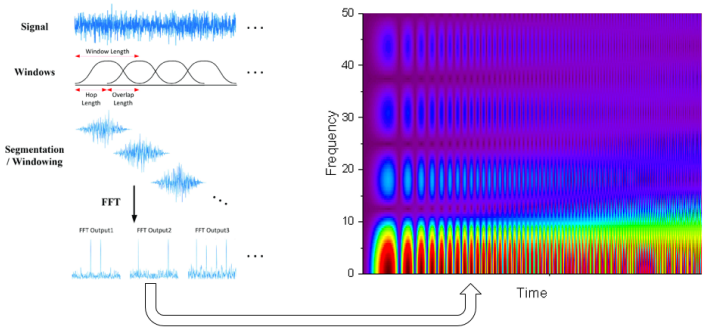
地把握人当前的情感状态，根据情感状态做出回应，可以极大地提升用户对AI产品的体验。这在商品推荐，舆论监控，人机对话等方面都有着重要的意义。例如，在销售过程中，了解用户对商品的满意度，可以帮助平台制定更好的销售策略；在影视行业，了解观众对节目的喜怒哀乐，能帮助制定更精彩的剧情以及安排特定节目的上线时间；在人机对话中，掌握人的情感状态可以帮助智能机器人做出恰当的回复，并适时地表达安抚和谅解，提升用户体验；在舆论方面，行政部门通过了解群众对热门事件的情感倾向、掌握舆论导向，从而更及时有效的进行舆情监控，为制定政策提供支持。情感识别还能应用于许多现实的场景中。情感识别算法具有很高的研究价值。

考虑到采集难度、隐私等因素，本文的工作聚焦于使用语音来识别说话人情感的语音情感识别（Speech Emotion Recognition, SER）任务。

2. 语音情感识别技术介绍

语音是日常生活中交流的主要媒介，它不仅传达了思想，还表达了说话人的情感状态。语音情感识别的目标是从语音中识别出人类的情感状态。其主要包含两个步骤：特征提取与分类器构建。

音频信号输入是近似连续的数值。提取音频特征通常首先对音频进行分帧，加窗，进行短时傅里叶变换（STFT）。然后得到了维度为$T \times D$的频谱特征，其中T表示帧数与时间长度相关，D是特征维度，每个维度对应不同的频率。有一些工作也会对此频谱进行一些mel滤波操作。



频谱特征包含丰富的信息，比如说话内容、节奏、语气、语调等等。与情感相关的语音特征提取仍然是一个尚未成熟研究方向。深度学习的出现简化了人工特征提出过程，使用数据驱动的方法，利用情感标签作为监督信号来训练深度模型提取与情感相关的隐含语义特征。由于音频输入

ectedness Index数据集

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发

【模型评估(一)】AI市场模型评估端到端流程打通

基于sentinel 1 和 2 数据的逐月Kmeans聚类分类的NDVI逐月统计

GEE ——多种机器学习方法（随机森林、cart、svm等）进行土地分类用光...

垃圾分类之场景化AI体验【玩转华为云】

ModelArts域适应算法EfficientMixGVB

对cann的算子编译注册工程的说明解析【对应工程1_custom_op】

AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...

2022年年终总结：日出万物生，日落满天星

GEE：社会联系指数（SCI）Social Connectedness Index数据集

CNN+Attention的方法。

传统的机器学习方法可以基于人工语音特征或者深度语音特征构建分类器，例如高斯混合模型（GMM），隐马尔科夫模型（HMM）,支持向量机（SVM）等经典方法。此外，得益于深度学习的发展，基于神经网络的分类器可以与深度特征提取器一起端到端（end-to-end）训练，得到情感分类器。

3. 语音情感识别面临的挑战

我们前面介绍了语音情感分析中常用的方法，但语音情感识别在实际中也面临着一些挑战：

1. 情感主观性与模糊性问题：语音情感识别是一个比较年轻的领域，在情感定义上缺乏官方标准。不同听者对同一段语音的情感可能有不同的观点。此外，一段语音往往有情感变化，主观性较强，导致许多研究工作没有普适性。

2. 情感特征提取和选择问题：语音说话人各种各样，情感类别多变，语音片段长短不一等，这些问题导致人工设计特征无法涵盖全部情感信息。另一方面，深度特征虽然效果好，但不具有可解释性。

3. 标注数据缺乏问题：深度学习方法取得很好的性能要求大量的高质量的标注数据。由于情感的主观性与模糊性，标注语音情感非常费时费力，同时要求大量专业人员。收集大量情感标注数据，是语音情感识别领域亟需解决的问题。

4. 如何解决数据缺乏的问题?

数据是深度学习的驱动力，大规模高质量的数据是深度学习取得成功的关键。然而，在很多实际问题中，由于标注代价问题，只存在少量的标注数据，这严重限制深度学习方法的发展。随着互联网社交平台的发展，每天都回生产大量的多媒体数据，大规模无标注的数据很容易获得。这就促进了能同时使用标注数据和无标注数据的半监督学习（Semi-Supervised Learning）方法的发展。另一方面，多媒体数据通常情况下都包含多个模态，因此也有一些工作探索利用一个模态的标注知识去加强在另一个模态上的任务的效果。下面介绍这两种方法。

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发

【模型评估(一)】AI市场模型评估端到端流程打通

基于sentinel 1 和 2 数据的逐月Kmeans聚类分类的NDVI逐月统计

GEE ——多种机器学习方法（随机森林、cart、svm等）进行土地分类用光...

垃圾分类之场景化AI体验【玩转华为云】

ModelArts域适应算法EfficientMixGVB

对cann的算子编译注册工程的说明解析【对应工程1_custom_op】

AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...

2022年年终总结：日出万物生，日落满天星

GEE：社会联系指数（SCI）Social Connectedness Index数据集

集，一个大规模的无标注数据集。其目的是利用无标注数据来增强，监督学习的效果。经典半监督学习方法包含很多类别，例如 self-training(自训练算法)，generative models（生成模型），SVMs（半监督支持向量机），graph-basedmethods（图论方法），multiview learing（多视角算法）等等。下面介绍几类主要半监督学习方法。

- 简单自训练算法（self-training）
  self-training算法的步骤为：（1）首先利用标注训练集数据训练分类器；（2）利用分类器对无标注数据进行分类，并计算误差；（3）选择分类结果中误差较小的样本，将分类结果作为其标签，加入到训练集。循环次训练过程，直到所有的无标注数据被标注。
- 多视角学习（multiview learing）
  这是self-training算法的一种。其假设每个数据可以从不同的角度进行分类。算法步骤如下：（1）在角度用标注数据集训练出不同的分类器；（2）用这些分类器从不同的角度对无标注数据进行分类；（3）根据多个分类结果来选出可信的无标签样本加入训练集。循环前面的训练过程。此方法的优点是不同角度的预测结果可以相互补充，从而提高分类精度。
- 标签传播算法（Label Propagation Algorithm）
  标签传播算法是一种基于图的半监督算法，通过构造图结构来找无标签数据和有标签数据之间的关系，然后通过这个关系来进行标签传播。

在深度学习上的半监督学习方法，叫做半监督深度学习。半监督深度学习主要包括三类：Fine-tune; 基于深度学习的self-training算法；半监督的方式训练神经网络。

Fine-tune方式，利用无标签数据训练网络（重构自编码或基于伪标签训练），然后使用有标签数据在目标任务上进行微调。

基于深度学习方法的self-training, 基本的步骤：（1）利用有标注数据训练深度模型；（2）利用深度模型作为分类器或者利用深度特征对无标签数据进行分类；（3）选择执行度高的加入有标签训练集，重复此过程。

半监督的方法训练深度网络包含许多技术，例如Pseudo-Label[1], Ladder Networks[2], Temporal Ensembling[3]，Mean teachers[4]还有FixMatch等等。

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发

【模型评估(一)】AI市场模型评估端到端流程打通

基于sentinel 1 和 2 数据的逐月Kmeans聚类分类的NDVI逐月统计

GEE ——多种机器学习方法（随机森林、cart、svm等）进行土地分类用光...

垃圾分类之场景化AI体验【玩转华为云】

ModelArts域适应算法EfficientMixGVB

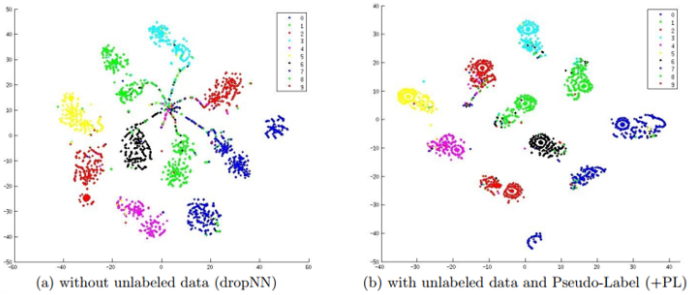
对cann的算子编译注册工程的说明解析【对应工程1_custom_op】

AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...

2022年年终总结：日出万物生，日落满天星

GEE：社会联系指数（SCI）Social Connectedness Index数据集

此方法将网络对无标签数据的预测结果，作为无标签数据的标签，来训练网络。方法虽然简单，效果却很好。从下图我们可以看出，加了无标签数据之后，同一个类别的数据点聚集得更笼了。



(a) without unlabeled data (dropNN) (b) with unlabeled data and Pseudo-Label (+PL)

Figure 1. t-SNE 2-D embedding of the network output of MNIST test data.

2. Temporal Ensembling[3]

   Temporal Ensembling是Pseudo-Label方法的发展。其目标是构造更好的伪标签。下图给出了此方法的结构图，此方法有两种不同的实现，即π-model和temporal ensembling。

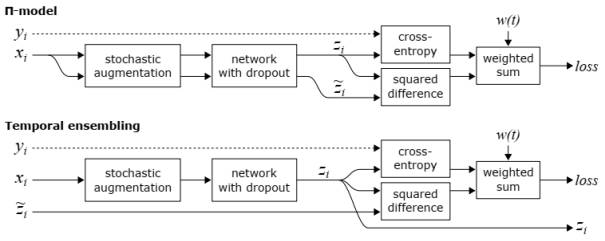


Figure 1: Structure of the training pass in our methods. Top: Π-model. Bottom: temporal ensembling. Labels y_i are available only for the labeled inputs, and the associated cross-entropy loss component is evaluated only for those.

   π-model的无监督代价是对同一个输入在不同的正则或数据增强的条件下模型输入应具有一致性，这样可以鼓励网络学习数据内部的不变性。

   Temporal ensembling对每一次迭代的预测z_i进行移动平均得个$\hat{z}_i$作为无监督训练的监督信号。

3. Mean teacher[4]

   Mean teacher方法另辟蹊径，从模型的角度提高伪标签质量，其奉行"平均的就是最好的"原则。对每次迭代之后的student模型参数进行移动平均（weight-averaged）得到teacher模型,然后用teacher模型来构造高质量的伪标签，来监督student模型的无标签loss。

4. FixMatch[5]

   FixMatch发扬了Temporal Ensembling方法中的一致性正则化（consistency regularization）原则，

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发

【模型评估(一)】AI市场模型评估端到端流程打通

基于sentinel 1 和 2 数据的逐月Kmeans聚类分类的NDVI逐月统计

GEE ——多种机器学习方法（随机森林、cart、svm等）进行土地分类用光...

垃圾分类之场景化AI体验【玩转华为云】

ModelArts域适应算法EfficientMixGVB

对cann的算子编译注册工程的说明解析【对应工程1_custom_op】

AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...

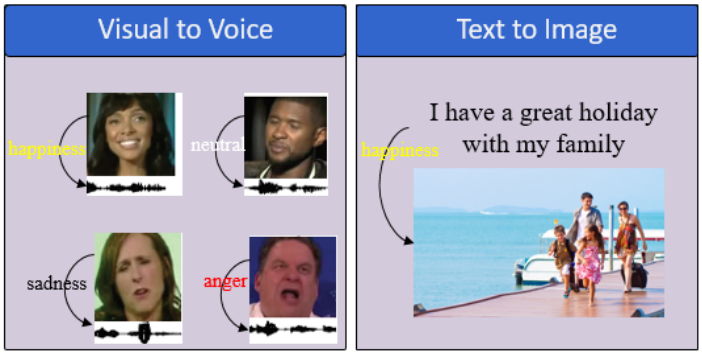
2022年年终总结：日出万物生，日落满天星

GEE：社会联系指数（SCI）Social Connectedness Index数据集

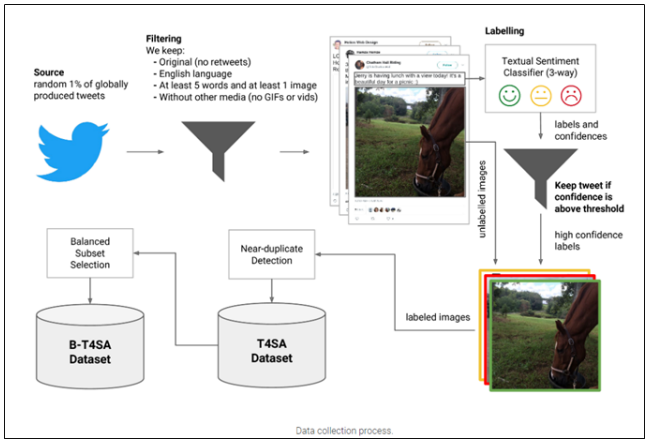
法利用弱增广的样本生成一个伪标签，利用此伪标签来监督模型对强增广样本的输出。

4.2 跨模态知识迁移

跨模态知识迁移基于多媒体数据中各个模态之间的内在联系，将标注信息由一个模态向目标模态迁移从而实现数据标注。如下图所示，跨模态知识迁移包括视觉到语音的迁移，文本到图像的迁移等等。下面介绍几种经典的跨模态知识迁移工作。



1. 基于跨媒体迁移的图像情感分析[6]
   此方法利用推特上成对的文本图像数据，完成图像情感分析任务，具体步骤如下图。



   其使用训练好的文本情感分类器，对文本进行情感分类，然后将标签直接给对应的图片。然后使用具有伪标注的图片训练图片情感分类器。

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发

【模型评估(一)】AI市场模型评估端到端流程打通

基于sentinel 1 和 2 数据的逐月Kmeans聚类分类的NDVI逐月统计

GEE ——多种机器学习方法（随机森林、cart、svm等）进行土地分类用光...

垃圾分类之场景化AI体验【玩转华为云】

ModelArts域适应算法EfficientMixGVB

对cann的算子编译注册工程的说明解析【对应工程1_custom_op】

AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...

2022年年终总结：日出万物生，日落满天星

GEE：社会联系指数（SCI）Social Connectedness Index数据集

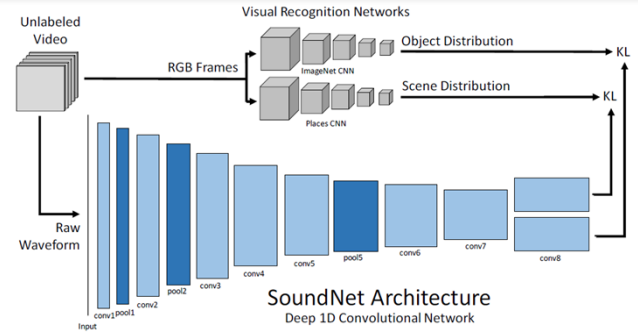
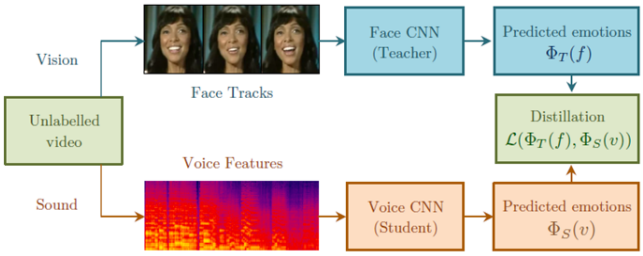


Figure 1: **SoundNet:** We propose a deep convolutional architecture for natural sound recognition. We train the network by transferring discriminative knowledge from visual recognition networks into sound networks. Our approach capitalizes on the synchronization of vision and sound in video.

通过预训练的视频对象和场景识别网络实现从视觉模态到语音模态的知识迁移，利用迁移的标签训练语音模型，完成语音场景或语音对象分类。

3. Emotion Recognition in Speech using Cross-Modal Transfer in the Wild[8]



此方法利用预训练好的人脸情感识别模型作为teacher模型，然后利用teacher模型的预测结果来训练语音情感识别模型。

5. 我们的语音情感识别方案

这一节将介绍我们处理标注数据缺乏的方案。

联合跨模态知识迁移与半监督学习方法

为了解决语音情感识别领域数据缺乏的问题，我们在2021年提出了联合跨模态知识迁移与半监督学习的架构，该方法在CH-SMIS以及IEMOCAP数据集上取得了语音情感识别任务当前最优的结果，同时我们将此工作发表在SCI一区期刊knowledge-based system上发表论文Combining cross-modal knowledge transfer and semi-supervised learning for speech emotion recognition。

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发

【模型评估(一)】AI市场模型评估端到端流程打通

基于sentinel 1 和 2 数据的逐月Kmeans聚类分类的NDVI逐月统计

GEE ——多种机器学习方法（随机森林、cart、svm等）进行土地分类用光...

垃圾分类之场景化AI体验【玩转华为云】

ModelArts域适应算法EfficientMixGVB

对cann的算子编译注册工程的说明解析【对应工程1_custom_op】

AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...

2022年年终总结：日出万物生，日落满天星

GEE：社会联系指数（SCI）Social Connectedness Index数据集

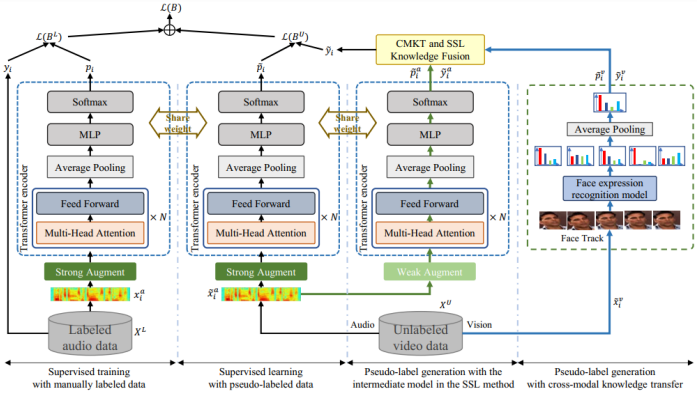


Figure 2: The high-level architecture of combining CMKT and SSL for SER.

我们的方案基于两个观察：

1. 直接跨模态标签迁移存在误差，因为人脸情感与语音语音情感之间的关系十分复杂，并不是完全一致。
2. 半监督学习方法，标注数据很少的情况下，表现并不好。模型的预测错误可能会不断的得到加强，导致模型在某些类别上精度很低。

我们的方法收到了多视角学习思路的启发，利用视频数据中存在两种模态，在两个模态上识别情感，融合它们获得更加准确的伪标签。为了进行语音情感识别，本方案首先提取了语音的STFT特征，然后进行了Specaugment数据增广。因为Transformer在建模序列数据的成功，本方案采用了Transformer的encoder进行语音的编码，最后利用均值池化来得到语音特征并分类情感。

跨模态知识迁移

为了进行跨模态情感迁移，本方案基于MobileNet模型利用大量的人脸表情数据集训练了一个性能强大的人脸表情识别模型。使用此模型对从视频中抽取的图片帧进行人脸表情识别。然后将多个帧识别的结果综合到一起得到整个视频段的人脸表情预测结果。

半监督语音情感识别

受到FixMatch中一致性正则化假设的启发，我们设计了半监督语音情感识别方法。具体的，此方法对语音样本输入采取了两种类型的增广，利用强增广方法SpecAugment算法获得到语音严重扭曲版频谱特征，利用弱增广方法（特征上的dropout等）得到变化不大的语音特征。模型使用弱增广的样本生成伪标签，来监督强增广的样本的训练。

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发

【模型评估(一)】AI市场模型评估端到端流程打通

基于sentinel 1 和 2 数据的逐月Kmeans聚类分类的NDVI逐月统计

GEE ——多种机器学习方法（随机森林、cart、svm等）进行土地分类用光...

垃圾分类之场景化AI体验【玩转华为云】

ModelArts域适应算法EfficientMixGVB

对cann的算子编译注册工程的说明解析【对应工程1_custom_op】

AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...

2022年年终总结：日出万物生，日落满天星

GEE：社会联系指数（SCI）Social Connectedness Index数据集

在模型的每一次迭代中，本方法利用弱增广样本生成一个伪标签，然后将其与跨模态迁移的伪标签进行融合，以提高伪标签的质量。本工作探索了两种融合方法，一个是加权求和，一个是多视角一致性。得到高质量的伪标签之后，用此标签监督强增广样本的训练。

模型通过多次迭代，不断提升伪标签质量。

相对于半监督学习方法和跨模态方法， 本方法在CH-SIMS和IEMOCAP数据集上均取得了最好的效果。结果如下：

Table 3: Results for 3-class speech sentiment analysis on the CH-SIMS dataset with data of iQIYI-VID as unlabeled data.

Model	F1(%)	WA(%)	UA(%)
Supervised Learning	42.94	45.51	44.91
Semi-supervised Learning			
FixMatch [10]	37.90	42.01	35.67
π-modal [9]	45.88	50.18	38.81
Temporal ensembling [9]	41.96	43.91	41.92
Direct Cross-modal Knowledge Transfer			
DCMKT [13]	36.58	36.76	44.94
DCMKT with fine-tuning	48.56	48.72	45.52
SSL & CMKT			
Consistent & Random $\epsilon = 0.5$ (ours)	**51.39**	**51.72**	**49.53**
Weighted Fusion $\alpha = 0.2$ (ours)	49.34	50.69	44.50

Table 5: Results for SER on the IEMOCAP dataset with the EmoVoxCeleb dataset as unlabeled data.

Model	F1(%)	WA(%)	UA(%)
Supervised Learning	57.40	58.42	59.43
Semi-supervised Learning			
FixMatch [10]	56.99	57.09	57.71
π-modal [9]	59.50	60.21	61.57
Temporal ensembling [9]	57.33	57.67	58.46
Direct Cross-modal Knowledge Transfer			
DCMKT [13]	24.15	28.72	27.99
DCMKT with fine-tuning	54.62	54.55	54.97
SSL & CMKT			
Consistent and random $\epsilon = 0.5$ (ours)	60.07	60.30	61.20
Weighted fusion $\alpha = 0.2$ (ours)	**61.06**	**61.16**	**62.50**

参考文献

[1] Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks
[2] Semi-Supervised Learning with Ladder Networks
[3] Temporal Ensembling for Semi-supervised Learning
[4] Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results
[5] FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发

【模型评估(一)】AI市场模型评估端到端流程打通

基于sentinel 1 和 2 数据的逐月Kmeans聚类分类的NDVI逐月统计

GEE ——多种机器学习方法（随机森林、cart、svm等）进行土地分类用光...

垃圾分类之场景化AI体验【玩转华为云】

ModelArts域适应算法EfficientMixGVB

对cann的算子编译注册工程的说明解析【对应工程1_custom_op】

AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...

2022年年终总结：日出万物生，日落满天星

GEE：社会联系指数（SCI）Social Connectedness Index数据集

Analysis in the Wild
[7] SoundNet : Learning Sound Representations from Unlabeled Video
[8] Emotion Recognition in Speech using Cross-Modal Transfer in the Wild

人工智能 AI平台

点赞 收藏 关注作者

热门文章

云推官之路的薪火相传【华为云云推官】
【云图说】第3期 初识虚拟私有云
【限时免费下载】华为云云享书库60+本精选电子书，...
【云小课合集】华为云小课最全合集来了，让您上云无...
GaussDB(for Redis)资料导航

评论（0）

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发
【模型评估(一)】AI市场模型评估端到端流程打通
基于sentinel 1 和 2 数据的逐月Kmeans聚类分类的NDVI逐月统计
GEE ——多种机器学习方法（随机森林、cart、svm等）进行土地分类用光...
垃圾分类之场景化AI体验【玩转华为云】
ModelArts域适应算法EfficientMixGVB
对cann的算子编译注册工程的说明解析【对应工程1_custom_op】
AI市场强化学习预置算法实践----使用强化学习训练智能体玩转Atari小游戏...
2022年年终总结：日出万物生，日落满天星
GEE：社会联系指数（SCI）Social Connectedness Index数据集

登录后可评论，请 登录 或 注册

推荐阅读

【云驻共创】如何使用华为云EI产品做创新应用开发

【模型评估(一)】AI市场模型评估端到端流程打通