

基于 CNN 的普米语孤立词语谱图分类

董华珍

黔东南州广播电视大学 人事科, 贵州 兴义 562400

摘要: 为实现普米语孤立词语谱图的分类, 引入基于卷积神经网络的语谱图模型, 该模型可以无监督学习语谱图特征实现分类. 本文搭建了一个 9 层的卷积神经网络模型, 利用彩色语谱图样本集进行训练, 并针对已训练好的模型, 通过实验检验各项因素对分类的影响, 从而得到适当的参数. 参数确定后, 进行卷积神经网络与支持向量机、BP 神经网络的对比实验, 验证算法的可行性和有效性. 实验显示基于卷积神经网络的普米语孤立词语谱图分类准确率达到 91%~95%, 这说明该算法是可行和有效的. 与支持向量机、BP 神经网络相比, 卷积神经网络具有自动提取特征, 避免过拟合问题, 适合大样本数据进行训练的优点.

关键词: 普米语孤立词; 语谱图; 分类; 卷积神经网络

中图分类号: TN912.34

文献标志码: A

文章编号: 1673-9868(2021)02-0160-09

近年来国内语音识别技术快速发展, 并取得显著成效. 国内语音识别领域处于领先地位的科大讯飞、百度语音实验室等对汉语语音识别率可以达到 95% 以上, 但对濒危少数民族语言, 如普米语、佤语等语言的语音识别技术研究仍处于空白状态. 普米族主要居住于云南省和四川省, 普米语为其民族语言, 由于没有文字, 语言和文化仅限于口口相传^[1]. 随着年长的人相继离世, 普米族的语言和文化正逐渐消亡. 为了促进普米族语言和文化的传承和发展, 语音识别技术介入迫在眉睫.

尽管目前语音识别的方法很多, 但总的来说可以归结为基于语音信号和基于语谱图的 2 种语音识别方法. 语音识别中, 同一个词不同说话者的语速不同, 同一个人发同一个音, 在不同时刻, 所用的时间也不相同^[2-3]. 因此, 通过语音信号来统计人类语音的发音特征非常困难. 基于语音信号的语音识别中提取的特征参数主要有梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficient, MFCC)、线性预测倒谱系数(Linear Prediction Cepstrum Coefficient, LPCC)等, 其中 MFCC、LPCC 是最有效的^[4], 但这 2 个特征参数的提取是比较困难的. 而基于语谱图的方法能够包含发音特点, 并将语音信号的所有特征以图像的形式显示出来. 通过傅里叶变换将语音信号转换为语谱图, 借助图像处理的方法提取语谱图特征, 最后利用分类器, 可以实现基于语谱图的语音识别, 从而大大减小了语音识别的难度. 近年来国内对基于语谱图的语音识别展开了研究. 如: 宋洋^[5]提取语谱图的边界特征和二值特征作为语谱特征, 通过构造 BP 神经网络实现维吾尔语音素分类. 唐闰臣等^[6]通过提取语谱图中的 Hu 不变矩特征、纹理特征和部分语谱特征, 基于 SVM 实现了对语音情感的分类. 梁士利等^[7]将语谱图的频域图矩阵进行投影后的值作为特征值, 并利用 SVM 实现二字汉语词汇语音识别.

基于语谱图的语音识别, 其本质是实现语谱图的分类. 普米语孤立词语谱图分类属于模式识别的范畴^[8], 模式识别算法的选择是普米语孤立词语谱图分类的难点也是核心问题. 目前常用的模式识别算法有

神经网络算法 (Neural Network Algorithm, NNA)、K-近邻分类器 (K-Nearest Neighbor Classifier, KNNC)^[9]、支持向量机 (Support Vector Machine, SVM)、卷积神经网络 (Convolutional Neural Network, CNN) 等。神经网络算法、K-近邻分类器进行模型训练时需要大量训练样本。此外,神经网络算法^[10]基于经验风险最小化原则,面临“过学习的风险”的问题。SVM 针对多分类和大样本数据时,单一核在学习过程中存在学习能力强泛化能力弱或泛化能力强学习能力弱等局限性,需要引入多核支持向量机^[11]。卷积神经网络^[12-14]是深度学习的一个重要算法,其提供了一种端到端的学习模型,通过梯度下降算法可对模型的参数进行训练,并自动地学习图像的特征完成图像的分类。

1 基于 CNN 的语谱图分类模型

基于 Google 公司开源的一个机器学习框架系统 Tensorflow,本文搭建了一个卷积神经网络模型,在此基础上利用语谱图样本集对模型进行训练。模型训练的步骤为:① 初始化普米语语谱图数据集;② 训练模型的参数;③ 模型的测试。由此得到语谱图分类的卷积神经网络模型。

1.1 CNN 模型的构造

CNN 模型由 7 个卷积层和 2 个全连接层组成。将彩色语谱图直接输入到卷积神经网络模型中,由于图像是彩色的,故模型通道数为 3。对输入的图片进行卷积操作后,采用 ReLU 激活函数,对卷积核的输出值进行非线性变换,接着采用 maxpoolig 完成最大池化操作。全连接层采用的激活函数是 ReLU 函数,为了防止模型训练过拟合,在全连接层后都添加 Dropout 层,设置 dropout_ratio 参数值为 0.5,即在模型训练过程中随机让网络中 50% 的节点不工作(输出置 0),表 1 列出了 CNN 网络结构的详细参数表。

表 1 CNN 网络结构参数表

网络层类别	输入层大小	滤波器	滤波器	步长	输出层大小
卷积层 1	200×150×3	64	1×1	1	200×150×64
池化层 1	200×150×64	—	2×2	2	100×75×64
卷积层 2	100×75×64	64	1×1	1	100×75×64
池化层 2	100×75×64	—	2×2	2	50×38×64
卷积层 3	50×38×64	128	1×1	1	50×38×128
池化层 3	50×38×128	—	2×2	2	25×19×128
卷积层 4	25×19×128	128	1×1	1	25×19×128
池化层 4	25×19×128	—	2×2	2	13×10×128
卷积层 5	13×10×128	256	1×1	1	13×10×256
池化层 5	13×10×256	—	2×2	2	7×5×256
卷积层 6	7×5×256	256	1×1	1	7×5×256
池化层 6	7×5×256	—	2×2	2	4×3×256
卷积层 7	4×3×256	256	1×1	1	4×3×256
池化层 7	4×3×256	—	2×2	2	2×2×256
全连接层 1	2×2×256	—	—	—	4 096
全连接层 2	4 096	—	—	—	1 650

1.2 CNN 模型的训练过程

1.2.1 数据的初始化

1) 在 Tensorflow 根目录下的子目录 Spectrogram-Final 文件夹中创建一个名为“tp-jpg”的文件夹,该文件夹用于存放语谱图样本集。部分训练样本如图 1 所示。

2) 在 Spectrogram-Final 文件夹中有一个名为“bmp2jpg”的运行脚本,可将 bmp 格式的图片转换为 jpg 格式,并利用 resize 函数对图像进行缩放。

3) 通过执行 Spectrogram-Final 文件夹中的 data-util.py 代码, 能将样本集严格地分为训练集和测试集, 并且得到图名称与标签.

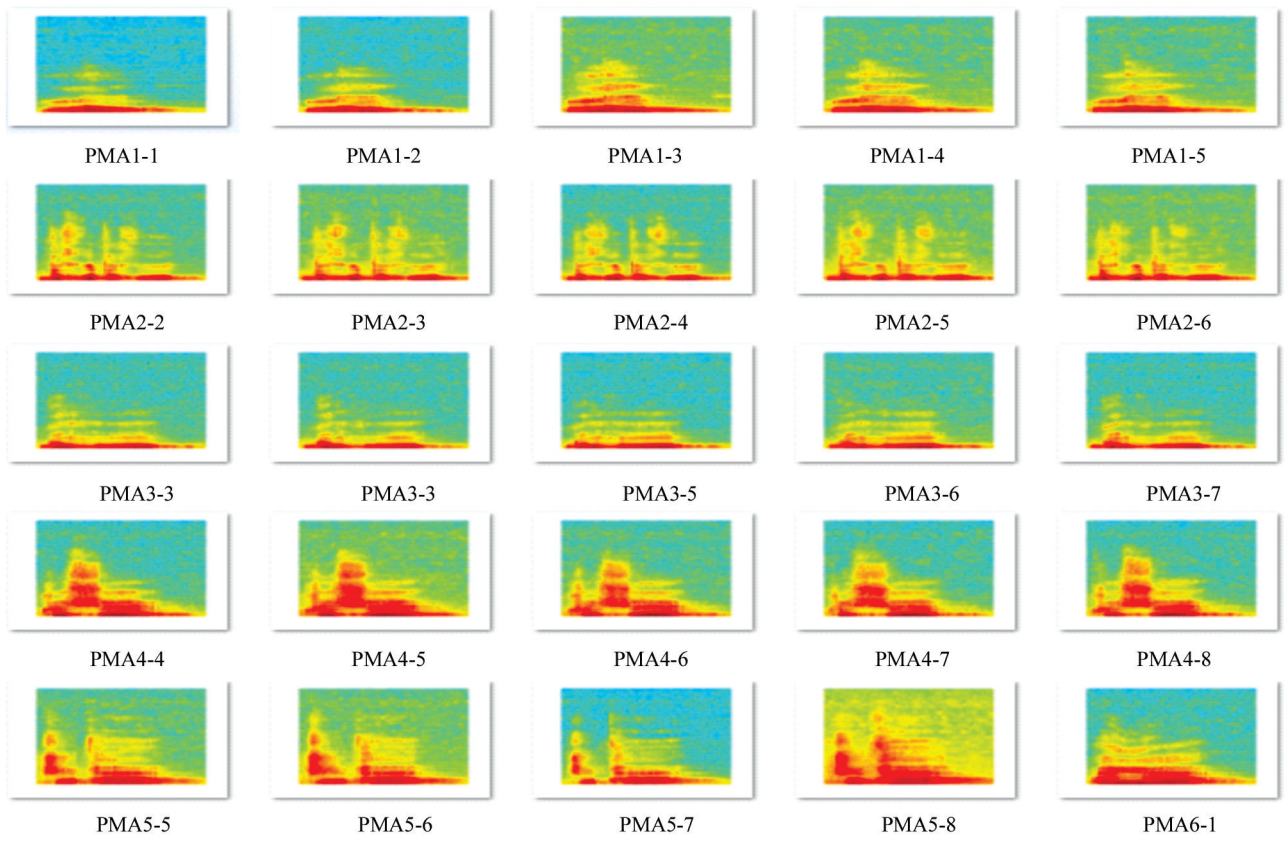


图 1 部分训练样本图

1. 2. 2 构造 CNN 模型

Spectrogram-Final 文件夹中的 cnn-model.py 为设计好的卷积神经网络模型脚本, 里面有卷积神经网络模型构造的详细内容, 包括每一个卷积层、池化层的参数设计. 设计好的卷积神经网络模型, 执行 cnn-model.py 的代码, 通过 Tensorboard 可以查看网络模型的示意图.

1. 2. 3 训练 CNN 模型

卷积神经网络的训练方式分为逐个样例训练方式、批量样例训练方式和随机批量样本训练 3 种方式. 本文设计的模型采用的是批量样例训练, 每个批次训练 150 个样本, 迭代 260 次, 在迭代过程中, 利用梯度下降算法和反向传播算法对模型的参数权值和偏置值进行修改.

1) Spectrogram-Final 文件夹中的 train.py 文件为卷积神经网络模型的训练代码. 将 data-util.py 所生成的训练样本和测试样本输入到上述已经设计好的模型中, 其中每批次训练的数量为 150 个样本, 迭代 260 个周期, 每迭代一次就测试一次数据. 网络的基础学习率为 0.0005, 随着网络的训练, 学习率将慢慢减小, 从而使结果收敛. 执行该程序, 即可训练上述定义的卷积神经网络模型.

2) 通过 Spectrogram-Final 文件夹中的 nohup.out GPU 文件, 可以查看模型训练过程中准确率的变化情况, 如图 2 所示, IFNO 代表分类准确率的变化情况. 此外, 也可以通过 Tensorboard-logdir train.log/命令启动 train.log 模型训练日志查看 Accuracy 和 cross entropy 的变化曲线, 从曲线中能够清楚地看到 Accuracy 和 cross entropy 的变化情况.

3) Spectrogram-Final 文件夹中的 checkpoint 文件夹, 用于存放训练好的模型, 可以自行查看训练后保存的模型、验证模型. checkpoint 文件夹中包含 checkpoint、model_260.ckpt.data-00000-of-00001、model_260.ckpt.index、model_260.ckpt.meta 4 个文件. checkpoint 是一个文本文件, 记录了训练过程中

所有中间节点上保存的模型的名称，图 3 为 checkpoint 文件中间节点上保存的模型名称示意图。model_260.ckpt.meta 文件是以“protocol buffer”的格式存储模型的结构图、定义的操作等信息，model_260.ckpt.data-00000-of-00001 文件保存网络结构的变量值，而 model_260.ckpt.index 文件则保存 model_260.ckpt.data-00000-of-00001 中数据和 model_260.ckpt.meta 文件之间的对应关系。

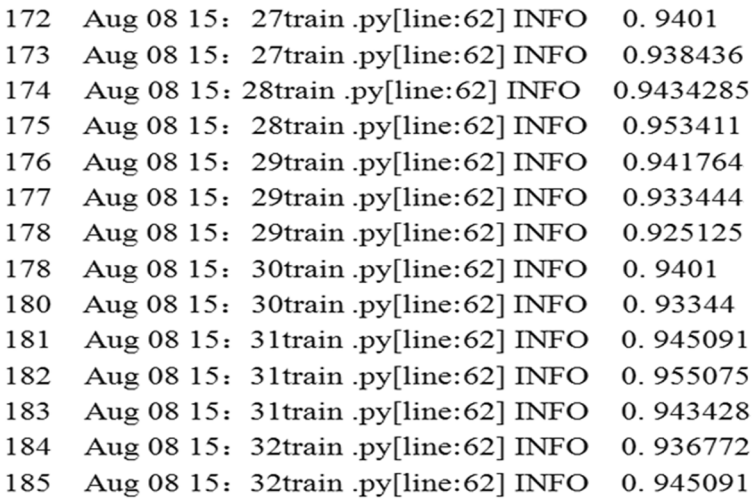


图 2 CNN 模型训练过程中准确率的变化情况示例图

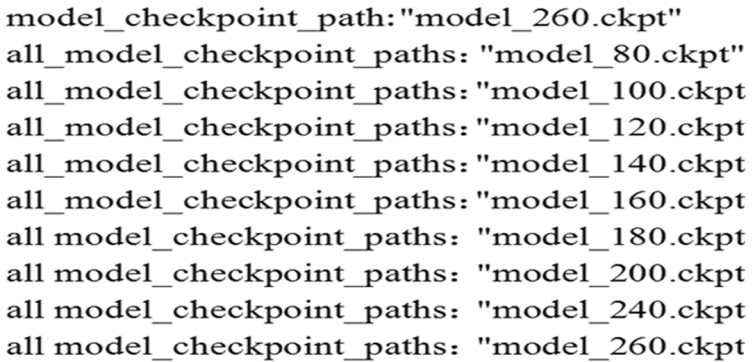


图 3 中间节点上保存的模型名称示意图

2 基于 CNN 的普米语孤立词语谱图分类实验

针对已训练好的卷积神经网络模型，通过实验检验各项因素对语谱图分类准确率的影响，从而得到卷积神经网络模型适当的参数。在参数确定后，将卷积神经网络与支持向量机(SVM)、BP 神经网络做了对比实验，以验证算法的可行性和有效性，并对 CNN 优于 SVM、BP 神经网络的原因进行分析。

2.1 数据集

实验室研究团队在云南省兰坪白族普米族自治县河西乡箐花村和录音棚中共采集了 1 650 个普米语孤立词语音语料，发音人为 2 男 2 女，每个词 8 遍，总计得到 52 800 条普米语孤立词语音。通过傅里叶变换可将这 52 800 条语音转换为语谱图，形成本文使用的样本集。

2.2 实验的评价标准

在测试实验中，当输入的一张语谱图经过卷积神经网络后，输出的类别与其对应的标签相符，则表明该语谱图的分类正确。反之，则网络对该语谱图的分类错误。将测试集输入至卷积神经网络中，最终的准确率作为该网络的评价指标，也作为普米语孤立词语谱图分类的最终准确率。准确率的高低作为卷积神经网络性能的评价指标。而且通过卷积神经网络与传统分类器进行分类效果的比较，可以判断卷积神经网络

性能的优劣. 实验的分类准确率定义为

分类准确率 = $\frac{\text{正确分类的语谱图数}}{\text{总的语谱图数}} \times 100\%$

2.3 CNN 的测试实验

实验包括彩色语谱图测试实验, 以及在不同学习率、不同样本比例、不同激活函数下的对比实验, 由此, 可以得到卷积神经网络模型的适当参数. 基于已设计好的 CNN 模型, 利用普米语孤立词语谱图进行训练, 并与传统分类器进行对比, 从而可以对 CNN 的性能进行判断.

普米语孤立词语音信号的采样频率为 44 100 Hz, 单通道, 语音信号为 .wav 格式. 傅里叶变换设置的点数为 1 024, 重叠的长度 512, 窗长为 1 024. 下述实验中从语料库中选取了 1 650 条普米语孤立词语音信号, 生成 13 200 张语谱图, 其中 11 550 张为训练集, 1 000 张为测试集, 650 张为验证集.

2.3.1 彩色语谱图实验

在 Python 程序中, 调用已经编写好的 Specgram 程序将语音信号转换为彩色语谱图, 生成的语谱图为 200×900 px. 利用 resize 函数可将语谱图大小修改为 200×150 px, 其修改后的语谱图直接作卷积神经网络的输入, 进行网络模型训练并测试性能. 图 4 为调用 Specgram 函数后, 语音信号转换为语谱图的示例, 图 5 为彩色语谱图 Accuracy 曲线图, 图 6 则为彩色语谱图 cross-entropy 曲线图.

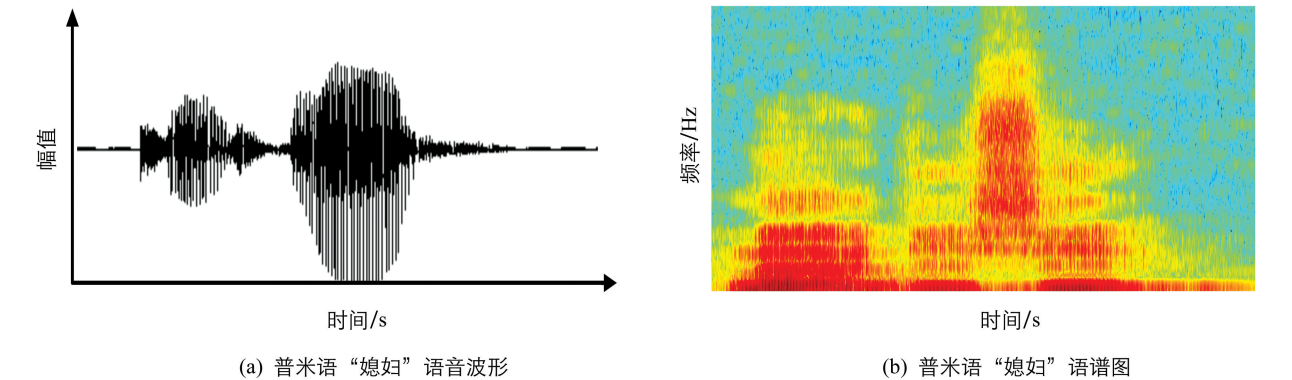


图 4 语音信号转换为语谱图的示例图

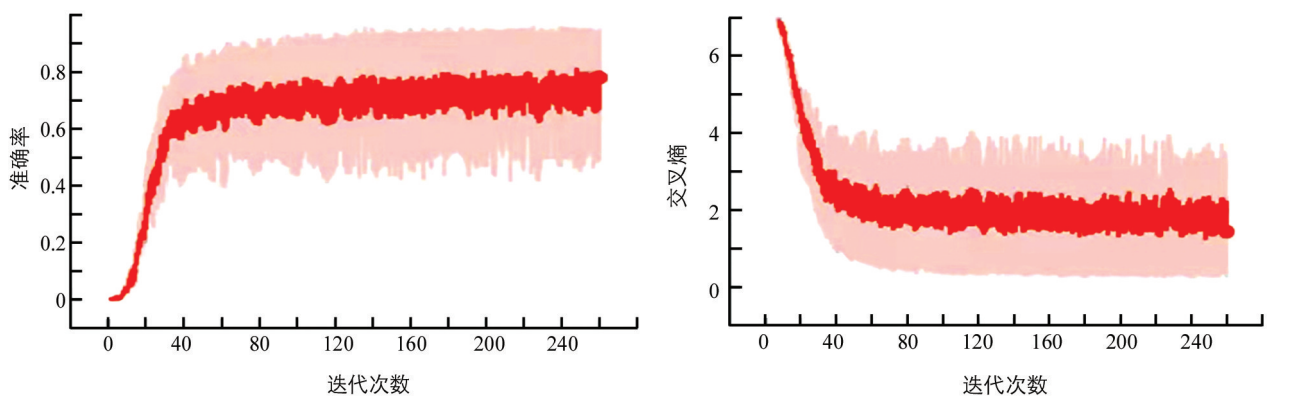


图 4 中能够看出, 语音信号转换为语谱图, 语音特征更加直观. 图 5 可观察到在卷积神经网络的训练过程中, 彩色语谱图的 Accuracy 值逐渐向 0.95 靠近, 并趋于平缓. 图 6 中 cross-entropy 的值逐渐向 0.01 靠近, 由此可以对网络模型参数进行调整.

2.3.2 学习率对比实验

卷积神经网络学习率的大小影响普米语孤立词语谱图的分类效果. 学习率过大, 卷积神经网络学习的

速度会加快，造成 Accuracy 曲线震荡或者发散；而学习率过小，会造成训练过程算法过早收敛。因此，本次实验采用 0.000 5、0.001、0.005、0.05、0.01、0.1 这 6 种不同的学习率进行对比实验，找出最佳学习率的值。表 2 为 6 种不同学习率下的对比实验结果。

表 2 不同学习率对比实验

学习率	迭 代 次 数						
	20	60	100	140	180	220	260
0.0005	71.56%	81.32%	85.76%	88.91%	91.09%	93.89%	95.01%
0.001	69.78%	69.78%	69.78%	69.78%	69.78%	69.78%	69.09%
0.005	69.78%	66.56%	69.09%	69.09%	71.73%	73.92%	73.92%
0.05	68.89%	64.87%	69.07%	70.98%	73.09%	74.56%	74.56%
0.01	68.84%	65.95%	70.98%	73.45%	74.56%	74.56%	74.56%
0.1	35.68%	35.08%	32.33%	35.86%	35.86%	35.86%	35.86%

从表 2 的对比实验中可以看出，学习率不同，语谱图分类准确率也不同。从表中还可以看出，虽然迭代次数在增加，但卷积神经网络学习率的值变化趋势越来越平缓，说明模型的学习能力趋于稳定。当学习率为 0.000 5 时，分类准确率最高，故本文所采用的卷积神经网络学习率为 0.000 5。

2.3.3 样本对比实验

表 3 为 4 种不同样本比例的对照表，表 4 则为每一类语谱图的 4 种样本比例对照表，表 5 和表 6 为不同规模样本集下的对比实验。从表 5 中可以得出少量的样本无法训练模型，对语谱图的最终分类准确率产生的影响较大。当样本比例为 7：1 时，分类准确率最高，达到 95.31%。

在实验中，语谱图共有 1 650 类，每一类有 8 张。每一类的训练样本集与测试样本集比例的不同，都会影响每一类的分类准确率。由此每一类训练样本集与测试样本集比例分别为 3：5、4：4、6：2、7：1，表 6 则为每一类不同样本比例的对比实验结果。表 6 可以得出结论，当每类样本的训练集与测试集之间的比率为 7：1 时，语谱图的分类准确率最高，约为 95%。样本比例为 6：2 时，准确率达到 78.98%，而样本比例分别为 3：5、4：4 时，分类准确率分别为 37%、44.49%。相比之下，当比例为 7：1 时，可在一定程度上提高最终的分类准确率。

表 3 4 种样本比例对照表

实验名称	样本数	训练集：测试集
实验 1	3 300	1：1
实验 2	6 600	3：1
实验 3	11 550	6：1
实验 4	13 200	7：1

表 4 每一类不同样本比例对照表

实验名称	训练集：测试集
实验 1	3：5
实验 2	4：4
实验 3	6：2
实验 4	7：1

表 5 4 种样本比例的对比实验结果

训练集： 测试集	迭 代 次 数						
	20	60	100	140	180	220	260
1：1	51.20%	55.60%	55.90%	45.80%	53.20%	53.20%	53.20%
3：1	57.56%	63.98%	67.98%	70.56%	69.45%	71%	71%
5：1	62.33%	63.32%	67.90%	73.35%	77.79%	83.09%	83.09%
7：1	63.39%	75.46%	89.94%	94.37%	95.31%	95.31%	95.31%

表 6 每一类不同样本比例的对比实验结果

训练集: 测试集	迭 代 次 数						
	20	60	100	140	180	220	260
3 : 5	23.12%	22.56%	26.89%	35.58%	37%	37%	37%
4 : 4	22.51%	35.78%	46.36%	33.56%	44.49%	44.49%	44.49%
6 : 2	61.76%	63.32%	74.90%	78.98%	78.98%	78.98%	78.98%
7 : 1	63.31%	75.23%	89.93%	94.38%	95.31%	95.31%	95.31%

2.3.4 全连接层输出节点对比实验

在卷积神经网络模型中,需要考虑不同的输出节点数是否会对模型的性能产生影响.由此,设置全连接层输出节点数分别为 1 024、2 048、4 096 进行对比实验.实验结果如表 7 所示.

2.3.5 激活函数对比实验

卷积操作实际上是一种线性操作,然而许多机器学习的问题是非线性的,需要将一个特征空间的向量通过非线性变换映射到另一个空间中,才能实现线性可分.激活函数是非线性变换的一种手段.常用的 4 种激活函数特点不一样,有的激活函数可能会在模型训练过程中出现“梯度消失”的问题,而有的则不会出现此问题.由此,需要通过实验找出分类效果好、收敛速度快的激活函数,实验结果如表 8 所示.

表 7 全连接层不同输出节点的对比实验

输出节点数	分类准确率/%
1 024	75~94
2 048	80~94
4 096	90~94

表 8 4 种激活函数对比实验

激活函数	训练时间/h	准确率/%
Softmax	28	78
Softplus	35	67
ReLU	18	95
Sigmoid	26	90

从表 8 中可以看出,在卷积神经网络模型中,以 ReLU 函数作为激活函数,模型训练时间最短,并且语谱图的分类效果是最好的.故本文模型选用 ReLU 函数作为激活函数.

2.3.6 CNN 模型参数设置

上述实验分别从输入的样本比例、学习率等方面进行对比实验,由此得到 CNN 模型的最佳设置参数,表 9 为卷积神经网络模型的最终参数设置.

2.4 基于 SVM、BP 神经网络和 CNN 的语谱图分类实验

为了说明该 CNN 模型对普米语孤立词语谱图分类具有可行性和有效性,在相同实验条件下,本文进行基于 SVM、BP 神经网络和 CNN 的对比实验.从普米语孤立词语谱图样本集中选择 400 张进行基于 SVM、BP 神经网络的分类实验,其中 320 张作为训练集,80 张作为测试集,提取每张语谱图的二值特征,得到 16×16 的特征矩阵.

1) 基于 SVM 的语谱图分类实验中,核函数与参数的确定将直接影响分类精度和泛化性能.针对这些问题,首先进行了线性核、多项式核、高斯径向基核、多层感知器核函数的对比实验,由实验结果得出:采用径向基函数作为 SVM 的核函数,预测分类准确率最高.其次,为了确定各个独立核的惩罚参数 c 和核函数参数 g 的值,先利用网格法来找最优参数.然而,该方法只能找到局部最优参数.免疫遗传算法将免疫算子添加到遗传算法中,避免了传统遗传算法不成熟收敛的问题,具有随机并行搜索的优点.因此,利用它进行 SVM 参数的全局优化.结果表明,当 $c=5.278$, $g=0.062\ 5$ 时,分类准确率最高,其结果如表 10 所示.

2) 基于 BP 神经网络的语谱图分类实验中,本文构建由输入层、隐含层、输出层组成的三层的 BP 神经网络模型进行语谱图的分类实验,模型的结构为 256-28-80,即输入层有 256 个节点,隐含层有 28 个节点,输出层有 80 个节点,其模型的基本参数学习率、期望精度、迭代次数分别为 0.000 5、 10^{-5} 、60 次.

3) 在上述的对比实验中, 得到卷积神经网络模型适当的参数, 并利用调整参数之后的模型进行普米语孤立词语谱图的分类实验. 3 种方法的分类准确率如表 10 所示.

表 9 CNN 模型最终参数设置		表 10 SVM、BP 神经网络和 CNN 的性能比较	
参数类型	模型参数设置	算法	准确率/%
网络结构	搭建 7 个卷积层、2 个全连接层的卷积神经网络	SVM	63
学习率	0.000 5	BP 神经网络	58
激活函数	ReLU 函数	卷积神经网络(CNN)	91
输出节点数	4 096		

在上述 SVM、BP 神经网络和 CNN 的对比实验中, 基于 SVM 的语谱图分类准确率为 63%, 而 BP 神经网络的分类准确率为 58%, 基于 CNN 的语谱图分类准确率最高, 达到 91%. 因此, 可以得出 CNN 与传统分类器 SVM 和 BP 神经网络相比, 语谱图分类准确率最高, 性能最好.

BP 神经网络在训练过程中容易出现过拟合的问题, 影响分类准确率. SVM 针对小样本, 虽然有很强的学习能力和泛化能力, 可得到现有信息下的最优分类模式, 但对多分类和大样本数据, 需要引入多核支持向量机, 核函数的组合以及参数寻优十分困难. 而卷积神经网络采用 ReLU 激活函数, 避免了过拟合的问题, 大样本数据则有利于模型参数的调整, 此外, CNN 能从图像中自动提取复杂的特征. 因此, 分类效果最好.

3 结 论

本文在机器学习框架系统 Tensorflow 上搭建了一个包含 7 个卷积层和 2 个全连接层的卷积神经网络模型, 并完成了普米语孤立词语谱图的分类实验. 通过不同样本比例、不同学习率、不同激活函数、不同输出节点数等的对比实验, 分析各项因素对分类准确率的影响, 得到模型适当的参数. 在此基础上, 分别对卷积神经网络、SVM 和 BP 神经网络的实验结果进行比较, 判断卷积神经网络模型的性能.

实验表明, 在基于卷积神经网络的普米语语谱图分类实验中, 分类准确率受学习率、激活函数、同一类训练样本与测试样本比例等因素的影响, 但通过模型的训练, 可找到适当的参数, 对模型参数进行调整. 相比于 SVM 和 BP 神经网络, 卷积神经网络的性能较好, 对普米语孤立词语谱图的分类是有效的. 通过实验分析各项因素对分类准确率的影响, 从而得到了适当的参数并在训练好的模型上进行普米语孤立词语谱图的分类实验.

参考文献:

[1] 胡文君, 傅美君, 潘文林. 基于 Kaldi 的普米语语音识别 [J]. 计算机工程, 2018, 44(1): 199-205.

[2] 余建潮, 张瑞林. 基于 MFCC 和 LPCC 的说话人识别 [J]. 计算机工程与设计, 2009, 30(5): 1189-1191.

[3] 李 原. 小词汇量连续语音识别系统的研究 [D]. 重庆: 西南大学, 2017.

[4] 游大涛. 基于听觉机理的鲁棒特征提取及在说话人识别中的应用 [D]. 哈尔滨: 哈尔滨工业大学, 2013.

[5] 宋 洋. 基于图像处理的维吾尔语音素语音学特征提取与应用 [D]. 新疆: 新疆大学, 2016.

[6] 唐闰臣, 冯月芹, 梁瑞宇, 等. 面向语音情感识别的语谱特征提取算法研究 [J]. 计算机工程与应用, 2016, 52(21): 152-156, 174.

[7] 梁士利, 魏 莹, 潘 迪, 等. 基于语谱图行投影的特定人二字汉语词汇识别 [J]. 吉林大学学报(工学版), 2017, 47(1): 294-300.

[8] 尹 恒, 付 佳, 何 凌, 等. 基于 PICGTFs 和 SSMC 增强的腭裂语音咽擦音自动识别方法: CN201810852508.2 [P]. 2019-2-1.

[9] CHEN S B, XU Y L, DING C H. Q, et al. A Nonnegative Locally Linear KNN Model for Image Recognition [J]. Pattern Recognition, 2018, 83: 78-90.

[10] 章鸣嫒, 陈 瑛, 沈 瑛, 等. 人工神经网络和支持向量机性能比较及其在 DMD 疾病识别中的应用 [J]. 上海理工大学学报, 2016, 38(4): 346-351.

[11] 王 霞, 董永权, 于 巧, 等. 结构化支持向量机研究综述 [J]. 计算机工程与应用, 2020, 56(17): 24-32.

[12] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks [J]. Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012, 1: 1097-1105.

[13] 李彦冬, 郝宗波, 雷 航. 卷积神经网络研究综述 [J]. 计算机应用, 2016, 36(9): 2508-2515, 2565.

[14] FUKUSHIMA K. Artificial Vision by Multi-layered Neural Networks: Neocognitron and Its Advances [J]. Neural Networks, 2013, 37: 103-119.

Primi Isolated Word Spectrogram Classification Based
on Convolutional Neural Network

DONG Hua-zhen

Personnel Section, Southwest Guizhou Autonomous Prefecture Radio & Television University, Xingyi, Guizhou 562400, China

Abstract: A spectrogram model based on convolutional neural networks is introduced to achieve the classification of Primi isolated words. This model can achieve classification through unsupervised learning of spectrogram features. A nine-layer convolutional neural network model is built, which is trained using the spectrogram sample set. For the trained model, the influence of various factors on the experimental results is tested by experiments, and thus the appropriate parameters are obtained. After the parameters are determined, a comparative experiment of convolutional neural network, support vector machine and BP neural network is performed, and the results show that the accuracy of Primi isolated word spectrogram classification based on convolutional neural network is as high as 91%~95%, thus indicating its effectiveness and feasibility. Compared with SVM and BP neural network, this convolutional neural network has the advantages of automatically extracting features, avoiding over-fitting, and adapting to large sample data for training.

Key words: Primi isolated word; spectrogram; classification; Convolutional Neural Network (CNN)

责任编辑 崔玉洁