

深度卷积神经网络在音乐风格识别中的应用

李 岳

赤峰学院, 中国·内蒙古 赤峰市 024000

【摘要】随着我国计算机信息技术、人工智能识别技术的不断发展, 各类新技术开始应用于音乐风格的识别中, 从目前的研究情况来看, 基于时间特征的音乐风格识别问题分类性能识别方面是存在着较大的问题的, 而卷积神经网络具有较强的信息捕捉能力, 因此将卷积神经网络技术应用在音乐风格识别过程中有着非常不错效果, 在本文的研究中, 结合行业内技术研究现状, 研究了深度卷积神经网络在音乐风格识别中的应用情况。

【关键词】深度; 音乐风格识别; 卷积神经网络; 应用

【课题】1. 2012.3主持赤峰市市政府科研课题 民族声乐艺术的文化价值研究 2. 2012.3参与赤峰市市政府科研课题 多元文化背景下赤峰地区高校教育研究 3. 2020.4参与 赤峰市蒙古族乐团在蒙古族文化传承中问题与对策研究 4. 2021年内蒙古自治区课题, 党史百年红歌的育人研究。

前言

在当前的音乐工作中, 对于音乐风格的识别与分类是一项非常重要的工作, 但是从音乐的特点来看, 音乐本身具有特殊性和复杂性, 而且音乐之间也没有明显地界限, 从这个角度来说自动分类音乐风格是当前技术中比较具有挑战性的一个问题, 从技术层面来分析, 音乐风格分类的主要技术难点就是要进行关键音乐信息的提取, 而当前的深度卷积神经网络技术则可以很好地实现这样的效果, 下文将结合技术特点进行研究。

1 深度卷积神经网络在音乐风格识别中的应用

1.1、Harmonic/Percussive 分离算法

音乐信号通常是由冲击声音和谐波声音所组成的, 两者之间有着鲜明的特性, 在深度卷积神经网络技术的应用过程中可以先使用Harmonic/Percussive分离算法将音乐信号中的内容进行分离, 在该技术应用中, 主要是根据谐波的特性展开的, 谐波的频率在时间方向是连续不断的, 冲击频谱在频率上也是连续的, 而采用这种分离算法分离之后可以使得谐波频谱在某个固定的频率上沿着时间轴进行连续平滑的分布, 从而更加有效地识别出各种音乐。

1.2 网络结构

在深度卷积神经网络结构的前基层都是作为特征提取器来对音乐的特征进行抓取的, 而深度卷积神经网络中最后一层可以通过 softmax 函数来进行特征的识别。具体来说, 深度卷积神经网络结构总共包含八层, 其中前五层是与 pooling 层交替的卷基层, 而剩下的三层则是用于分类的全连阶层, 在具体识别中, 该技术首先识别音乐信号, 然后转化为图谱, 将其输入到第一个卷积滤波器之中, 接着在深层网络结构中对输入的内容进行滤波处理, 进行响应归一化。在第三、第四和第五个卷基层中是相互连接的, 它们之间没有 pooling 层和归一化层, 经过这几层的处理之后, 最后一个全连接层输出的结果便是最终的识别分类结果。

1.3 网络训练与学习方法

在音乐风格的识别过程中, 使用的网络结构是一个分层的深度卷积神经网络, 在处理中先通过卷积输入内容, 然后使用核滤波器提取局部的特征, 在卷基层的处理过程中通过线性卷积滤波器以及非线性激活函数生成特征图, 具体来说, 就是在

同一层中神经元的输出形成一个平面, 我们将该平面称之为特征图, 然后利用 pooling 获得卷积特征图并且将其滤波到下一层之中, 接着使用 local receptive field 通过设置不同的核滤波器来获取不同的特征图。

1.4 系统结构与设计

在系统结构与设计过程中, 可以分两个层面继续, 首先要做好前台交互界面的设计, 在前端的页面上具有音频上传的功能, 在这里用户可以上传音乐, 为了使得该系统的应用效果更加优化, 在前端用户中还可以播放上传的音乐, 同时, 在播放的过程中页面会实时地展现出当前音乐的流派概率。后端的处理中可以采用 keras 作为深度学习框架并且用来搭建神经网络, 在训练中, 将 GTZAN 数据集解压到专门用于识别数据集的文件夹之中, 在该数据集之中, 每个音频都有着相应的标签和类别, 在训练开始的时候系统会先调用 librosa 工具, 利用该工具读取音频文件, 然后将音频文件转化为梅尔谱图, 接着将其转化为数据后才能够进行训练, 在训练完成之后保存神经网络即可实现整个过程。在后台的处理部分中, 主要是接收前台页面上上传的音频数据, 并且直接使用已经训练好的神经网络, 然后对音频数据进行分类处理, 最后将分类的结果返回到前端操作系统中, 向使用者展示出来。

2 结语

音乐风格的识别是一项技术难度较大的工作, 而使用深度卷积神经网络技术则可以完成这样的过程中, 该技术从目前的应用效果来看, 识别和判断能力在小数据库中有着比较广泛的应用, 在未来的工作中还需要我们扩充更多风格的音乐曲目来提高该技术的识别率。

参考文献:

- [1] 张秀, 李念祖, 李晓强, 王淞昕, 李会永, 李伟. 基于 Chroma-based BOW 特征的多版本音乐识别[J]. 小型微型计算机系统, 2015, 36 (02): 397-400.
- [2] 杨丽萍, 卢岭, 刘莉, 梁耕田, 金辉, 陶朵朵. 人工耳蜗使用者汉语声调感知与音乐感知相关性研究[J]. 中华耳科学杂志, 2019, 17 (06): 905-909.
- [3] 刘彪, 黄蓉蓉, 林和, 苏伟. 基于卷积神经网络的盲文音乐识别研究[J]. 智能系统学报, 2019, 14 (01): 186-193.