



西南财经大学
SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS

2022 届 本科毕业论文（设计）

论文题目： 一个个性化推荐系统——以网易云音乐为例

学生姓名： _____

所在学院： 经济信息工程学院

专 业： _____

学 号： _____

指导教师： _____

成 绩： _____

2022 年 3 月

西南财经大学

本科毕业论文原创性及知识产权声明

本人郑重声明：所呈交的毕业论文是本人在导师的指导下取得的成果，论文写作严格遵循学术规范。对本论文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。因本毕业论文引起的法律结果完全由本人承担。

本毕业论文成果归西南财经大学所有。

特此声明

毕业论文作者签名：

作者专业：

作者学号：

2022 年 03 月 20 日

摘要

ZetDard 推荐系统是一个基于网易云音乐的推荐系统。其基本功能有获取用户信息、用户歌单、歌曲信息、歌曲评论以及获取对用户进行歌曲、歌手推荐等功能。本文先通过 Node.JS 部署了网易云音乐的 API 到本地，通过数据爬虫链接 API 获取源数据。针对获得的音乐源文件和用户以及歌曲的数据通过 Python 进行解析并且存储到本地，接着通过 Networkx 库对数据进行解析和读取，在对数据进行清洗、标准化之后匹配不同类型的数据，使用自定义的关联分析算法计算歌曲节点之间的相互影响程度，再经过数据化处理后得到了邻接矩阵构建出歌曲之间的复杂网络。通过复杂网络节点相似算法为用户初步推荐交友对象、喜欢歌曲、喜欢歌手等服务，在此之后，通过分析特定用户的歌曲音频特征，运用 SVM、卷积神经网络、聚类等机器学习和深度学习模型为用户达成个性化推荐。并且使用了多模态技术——将歌曲歌词与歌曲的音频特征映射到一个特征向量中，发现多模态的使用有效的提升了模型效果，同时针对数据不足的情况，通过 DCGAN 模型模型进行数据增强，然后在进行模型的训练。经过多重模型的比对和训练后，本文得出的结论是：多种模态融合的判别效果显著好于单一模态信息，卷积神经网络模型的预测准确率显著高于常用的机器学习模型。

但是本文在构建的模型，也明显存在算力开销较大、所需时间过长等问题，故而本文对业界的推荐机制猜测，业界的音乐推荐算法应该使用图论或者聚类所需时间和算力开销较小的算法。

关键词：生成对抗网络；卷积神经网络；支持向量机；复杂网络；音频分析

Abstract

In this paper, the API of NetEase Cloud Music is deployed locally through Node.js, and source data is obtained through data crawler link API. For the music, and songs of data through the Python source files and users parsed and stored locally, and then through the Networkx library to parse and read data in the data cleaning, standardized matching different types of data, using custom correlation analysis algorithm to calculate the mutual influence between songs nodes, After data processing, adjacency matrix is obtained to construct the complex network between songs. The node similarity algorithm of complex network is used to preliminarily recommend friends, favorite songs, favorite singers and other services for users. After that, by analyzing the audio features of songs of specific users, machine learning and deep learning models such as SVM, convolutional neural network and clustering are used to achieve personalized recommendation for users. In addition, multi-modal technology was used to map song lyrics and song audio features into a feature vector, and it was found that the use of multi-modal effectively improved the model effect. At the same time, in view of insufficient data, data enhancement was carried out through DCGAN model, and then the model training was carried out. After comparison and training of multiple models, the conclusion of this paper is that the discrimination effect of multiple mode fusion is significantly better than that of single mode information, and the prediction accuracy of convolutional neural network model is significantly higher than that of commonly used machine learning model.

However, the model constructed in this paper also has obvious problems such as high computational cost and long computational cost. Therefore, we guess the recommendation mechanism of the industry, and the music recommendation

algorithm of the industry should use graph theory or clustering and other algorithms with low computational cost and time.

Keywords: Adversarial network generation; Convolutional neural network; Support vector machine; Complex network; Audio analysis

目录

1. 概论	4
1.1 背景.....	4
1.2 文献综述.....	5
2.数据获取.....	6
2.1API 数据.....	6
2.2 网页爬虫数据	7
3.推荐过程.....	8
3.1 复杂网络推荐算法	8
3.2 基于卷积神经网络的音频推荐算法	11
3.3 融合多态信息的卷积神经网络推荐算法	13
3.4 基于支持向量机的歌曲推荐	14
3.5 关于数据不平衡的补全方法——DCGAN	15
3.6 协同聚类.....	16
4.总结	17
参考文献	18
致谢.....	19

1. 概论

1.1 背景

随着互联网技术的普及和人工智能的发展，越来越多的行业实现了与大数据的紧密融合，并对以往数据进行收集和分析，针对用户特征进行定制项目和推荐，更大程度上实现了用户的留存。同时，推荐系统的产生也一定程度上缓解了信息过载的问题。

关于个性化推荐系统，维基百科定义为：推荐系统是一种信息过滤系统，用于预测用户对物品的“评分”或“偏好”。纵观个性化推荐系统的发展历史，发现这个概念的提出始于 1995 年，卡耐基·梅隆大学的 Robert Armstrong 等人在美国人工智能协会上提出了个性化导航系统 Web Watcher。1995 年，麻省理工学院的 Henry Lieberman 在国际人工智能联合大会上提出了个性化导航智能体 Litizia。1996 年，Yahoo 推出了个性化入口 My Yahoo。1997 年，AT&T 实验室提出了基于协同过滤的个性化推荐系统 PHOAKS 和 Referral Web。1999 年，德国 Dresden 技术大学二点 Tanja Joerding 实现了个性化电子商务原型系统 TELLIM。2000 年 NEC 研究院的 Kurt 等人为搜索引擎 CiteSeer 增加了个性化推荐系统。2001 年，纽约大学的 Gediminas Adoavicius 和 Alexander Tuzhilin 实现了个性化电子商务网站的用户建模系统 1: 1Pro。2001 年，IBM 公司在其电子商务平台 Websphere 中增加了个性化功能，以便开发个性化电子商务网站。直至 2009 年 7 月，国内首个个性化推荐系统科研团队北京百分点信息科技有限公司成立，国内的个性化推荐系统有了更深的研究。

网易云音乐作为后起的音乐软件，自发布以来，该款产品的发展势头可谓十分强劲。近年来，网易云音乐平台受到了越来越多的用户的认可，据网易云音乐在 2020 年发布的 7 周年内部邮件披露，网易云隐喻的用户量已经突破了 8 亿，平台

入驻原创音乐人已超 16 万。其中，用户年龄段主要集中在 90 后，目前吸收的用户群体有向 00 后偏向的趋势。而网易云音乐能够在短时间内获得如此快速增长，以及超高用户的口碑粘性，这一成功与其用户第一的产品观是分不开的。在用户至上理念的驱动下，网易云音乐重点打造了针对用户个性的推荐系统。同时网易云音乐拥有着部署便捷，文档清晰的 API 接口，因此本文选择了网易云音乐来作为个性化的推荐系统的数据来源。选择网易云这么一款较为成功的，并且同时是以其独特的歌曲推荐而闻名的音乐软件来探究推荐系统，也是富有现实意义的。音乐个性化推荐系统，可以将其理解为，基于用户历史的听歌数据，针对其歌曲风格偏好程度、歌手喜爱偏好程度、听歌时间等因素，进行自动化地相似歌曲或歌手的推荐。

1.2 文献综述

网易云的个性化推荐是其拥有较高用户留存度的主要原因，这背后离不开精准的推荐系统以及算法对音频、用户、歌手等信息的支撑分析。Markus Schedl 研究了国家集群和相应的音乐消费偏好原型，确定了原型国家集群，分析了这些聚类在音轨层面上对应用户的音乐偏好，以及每个聚类中国家的语言、历史和文化背景，运用 VAE 架构对用户进行了国家分类的推荐。Eva Zangerle 将音轨的音频特征和世界幸福报告的数据相结合，并添加了霍夫斯泰德文化维度的影响因素，实现了更高精度的推荐。

随着人们对体验感的要求越来越高以及人工智能技术应用的更加广泛，国内学者对于其研究也日益增多。陈长风学者基于 CNN-LSTM 组合模型，从音频信号中提取特征参数，对音频情感进行了分类，并通过支持向量机、卷积神经网络、循环神经网络、组合网络模型等提高了情感分类的准确性；学者龚志基于用户的协同过滤方法，计算出目标用户与其他用户的相似度，在用户社区中找到与目标用户最相似的部分用户，再有这部分用户对指定音乐的评分来预测目标用户对该资源的评分，并引入用户兴趣模型，通过 VSM（向量空间模型）对歌词进行提取和降维，进行音乐特征的提取，最终实现多模态融合，从而产生个性化推荐结果；李春阳等学者利用深度学习方法，应用 CNN 网络显著的特征提取能力和平移不变性对异音进行分类识别，还提出了 ADGAN 网络模型进行数据增强，增加了数据量和数据的多样性，提高了模型的性能；付炜和杨洋则采用了卷积神经网络和随机森林相结合的混合模型，将音频转换成谱频图，再利用卷积神经网络提取频谱图的高层特征，简化

了特征提取的复杂过程。

综上所述，国内外学者均对个性化推荐音乐进行了多角度的分析，并提高推荐的精确性。在个性化推荐的研究中，将本文的模型与原有研究成果进行了对比，选取了国内具有代表性的 APP——网易云音乐的数据，通过用户歌单中歌曲和歌词的特征，并基于音频波普特征和神经网络进行分析，抽取数据中的特征，得出用户的听歌偏好，从而实现相关歌曲、歌手、相似用户的推荐以及音频的生成。

2.数据获取

本组在数据获取中，主要运用了两种形式，一是通过网易云 API 直接按需进行数据的调取，例如用户、歌曲、歌词、评论、歌单等；二是运用 python 进行网页爬虫，实现歌曲相关信息与歌曲 ID 的一一对应关系，并通过动态 JS 加密爬取得到用户 id 和昵称的对应关系。

在数据获取时，本文采用了分布式爬取，不间断地进行数据爬取，同时，本文还把获取到的图片、音频等进行处理，将单调的数据可视化，提高了数据的可读性和延展性。

2.1API 数据

本文选择了网易云音乐 API (NeteaseCloudMusicApi)¹来作为本文的主要数据来源，此 API 是网易云的官方 API，支持 Vercel 部署，TypeScript，以及 Node.js 部署。可以通过数据接口提供给用户登录、评论歌曲、查看评论、获得每日推荐、分析歌曲详细信息和用户电台等多种数据。本文选择通过 Node.js 将该 API 部署在 Ubuntu 系统本地，然后通过 python 爬虫发送访问请求，获取到数据后进行 json 数据解析，然后获得了以下信息：

表 2-1 通过 API 获取的数据

获得的信息	包含	数量
用户个人信息	用户 ID，用户昵称	3000+

用户歌单信息	歌单名，歌单所包含的歌曲	15000+
歌曲评论信息	歌曲 ID，评论文本，评论用户	150000+
歌曲文件	歌曲本身 mp3wenjian	2000+
歌曲详细信息	演奏歌手	2000+
歌手	歌手所演奏的歌曲	3000+
好友信息	获得一个指定用户的好友	8000+

2.2 网页爬虫数据

通过调用网易云 API 接口，本文已经获得了相关信息，但是目前数据仍比较分散，id 与其他信息之间并没有实现一一对应。因此本文通过 python 进行爬虫，获取歌曲的相关信息，如歌曲 id、歌曲昵称、歌曲时间等，以及通过分析 json 加密参数，获取网易云网页上的用户的相关信息，包括用户 id、用户昵称、用户地区等信息，实现了歌曲 id 与歌曲信息之间的对应、用户 id 与用户信息的对应，提高了后续工作的灵活性，使得对数据的分析更加具有针对性。在网易云音乐的官网上输入用户昵称进行用户查找后，浏览器会接收到服务器端响应的查找结果，其中包含用户信息(id、昵称、所在地区等)。由于在网页中显示用户信息的部分是 js 动态加载的，因此本文无法直接通过 Requests 和 BeautifulSoup 进行数据的爬取，并且后续发现还需要进行参数加密处理。该部分爬虫具体实现过程如下：

- 在 chrome 浏览器打开开发者工具后，通过分析找到相应的资源文件；
- 分析资源文件的请求头信息，找到相应的 request url；
- 分析资源文件的发起方，找到对应的 js 文件；
- 通过分析 js 文件得到参数加密方式；
- 使用抓包工具替换 js 文件，得到加密前的初始参数；
- 通过 request url，并携带经过修改初始参数(修改用户昵称)，并使用相应加密方式后得到的数据，就能够获取包含用户信息的 json 格式的数据；
- 从获取到的 json 数据中提取相应用户的信息即可。

本文的所有数据已经存放在了项目随附的 GitHub 项目中。

3.推荐过程

3.1 复杂网络推荐算法

考虑到传统的推荐系统都是基于运筹学方法来加以推荐，所以本文也在这里基于复杂网络做出自己的推荐结果，本文通过数据采集（见部分三），获取了大量的歌曲、用户、歌手信息，通过这些信息本文构建了一张巨大的歌曲复杂网络，具体而言，构建一张具有三类节点的复杂网络。

这三类节点主要是：

- 用户节点，类型标注为“users”。
- 曲节点，类型标注为“music”。
- 歌手节点，类型标注为“artists”。

为了构建三种节点的关系，本文重新定义三种节点之间边的出现方式。

在用户节点和歌曲节点之间，如果歌曲出现在用户收藏的歌单里，则默认将用户节点和歌曲节点连接起来。而歌曲节点之间，本文首先为每一首歌曲计算出收藏了该歌曲的用户列表，然后让歌曲之间的用户列表两两取交集，然后交集的数量除以该歌曲的所有用户数量，即歌曲节点之间的网络是一个有向图关联网络，同时节点之间的边的权值一定会处于 0 到 1 之间。在歌曲节点和歌手节点之间，则在歌曲与其原唱歌手之间构建联系。

通过以上连接方式，本文的用户——歌曲——歌手网络构造完成，如图所示（图为可视化效果，因为只选择了很小一部分来作为可视化）。

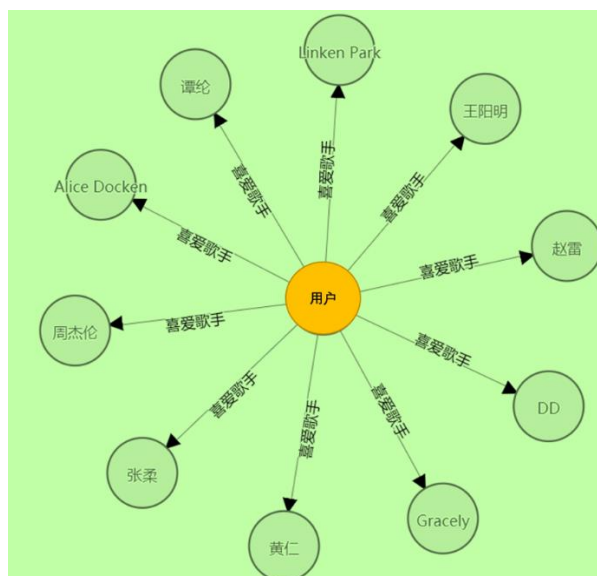


图 3-1 复杂网络可视化示例

在开始图节点算法之前，本文需要介绍一下聚类算法与变分自编码器。将物理或抽象对象的集合分成由类似的对象组成的多个类的过程被称为聚类。由聚类所生成的簇是一组数据对象的集合，这些对象与同一个簇中的对象彼此相似，与其他簇中的对象相异。聚类分析是由若干模式组成的，通常，模式是一个度量的向量，或者是多维空间中的一个点。聚类分析以相似性为基础，在一个聚类中的模式之间比不在同一聚类中的模式之间具有更多的相似性。本文将聚类分析应用于此处，是为了找出产业节点之间的内在相似性，通过将节点进行聚类，方便进行宏观政策调控和管理。

但是在应用聚类算法时，因为矩阵维度问题会造成算法运算时间过长，而本文中，因为要使用到大量的邻接矩阵数据，数据维度较高，分类效果较差，因此本文选用自编码器对数据做一个特征降维处理。

自编码是一种以重构输入信号为目标的神经网络。它是无监督学习领域中的一种，可以自动从无标注的数据中学习特征。

自编码器由三个神经网络层组成：输入层、隐藏层和输出层。其中，输入层的样本也会去充当输出层的标签角色，即这个神经网络就是一个尽可能地复现输入信号的网络。自编码神经网络要求输出尽可能等于输入，并且它的隐藏层必须满足一定的稀疏性，通过将隐藏层中后一层比前一层神经元数量少的方式来实现稀疏结果。这相当于隐藏层对输入进行了压缩，并在输入层中解压缩。整个过程中一定会丢失部分信息，但是训练能够使得丢失的信息尽量的少，最大化地保留其主要特征。

变分自编码神经网络学习地不再是样本的个体，而是样本的规律。通过增加一个约束项将编码器生成为服从高斯分布的数据集，然后按照高斯分布的均值和方差规则任意取相关得到数据，并且将数据输入编码器还原为样本。自编码器的训练过程如图所示，在进行 270 个 Epoch 之后，损失降低到了一个比较低的位置：

自编码器的训练过程如图所示，在进行 270 个 Epoch 之后，损失降低到了一个比较低的位置：

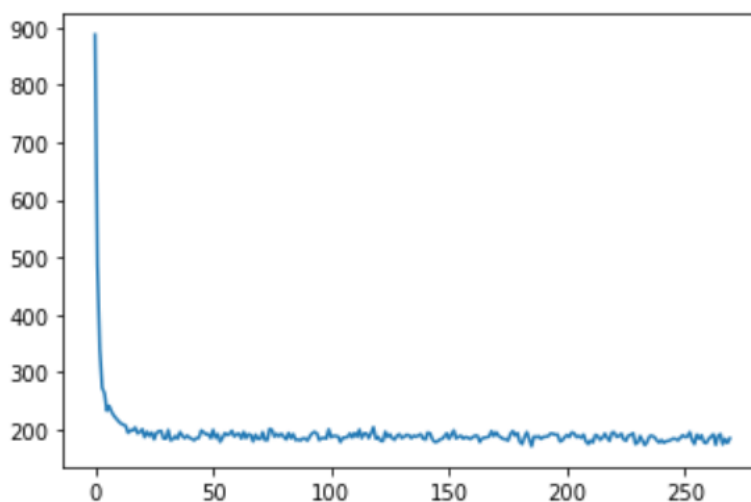


图 3-2 自编码器的训练结果

接下来运用图论算法来为用户进行初步的推荐：

- 度量复杂网络中用户节点的相似性，具体包含于：度量用户节点所连接歌曲节点的相似性、度量用户节点所连接节点的个数相似性，为了加快运算速度，此步中用户由该用户歌单中的歌曲所连接的所有用户节点组成（相当于，只有收藏了该用户的歌单中的歌曲的用户才会被纳入相似性比对），在计算完成相似性之后，再设立阈值，并且进行排序，得出该用户的推荐用户。
- 在得到相似用户后，本文将相似用户的歌单中的歌曲做合集，然后与用户歌单做差集，并且按照歌曲出现频率进行排序，然后得到对用户的歌曲推荐结果。
- 同时，本文考量已经构建好的歌曲节点网络，使用图自编码器为歌曲节点编码，然后做歌曲节点的聚类，然后选择与当前用户歌单中的歌曲较为相似的歌曲节点作为推荐结果推送给用户。
- 在得到相似歌曲后，本文查找该歌曲的创作歌手，并且对比用户歌单中的

歌曲中的歌手，然后推荐给用户作为歌手推荐的结果。

在此，本文得到了基于图论的用户推荐机制。

3.2 基于卷积神经网络的音频推荐算法

人工神经网络算法是由生物学中的大脑神经网络启发，为深度学习的一个核心。基于他的多变以及强大的模型，让神经网络算法成为一个解决例如分类大量图片等复杂机器学习问题的理想算法。抽象于大脑的神经网络模型，简单的神经网络模型有三层结构:输入层、隐藏层和输出层。如下图就是一个非常简单的神经网络模型：

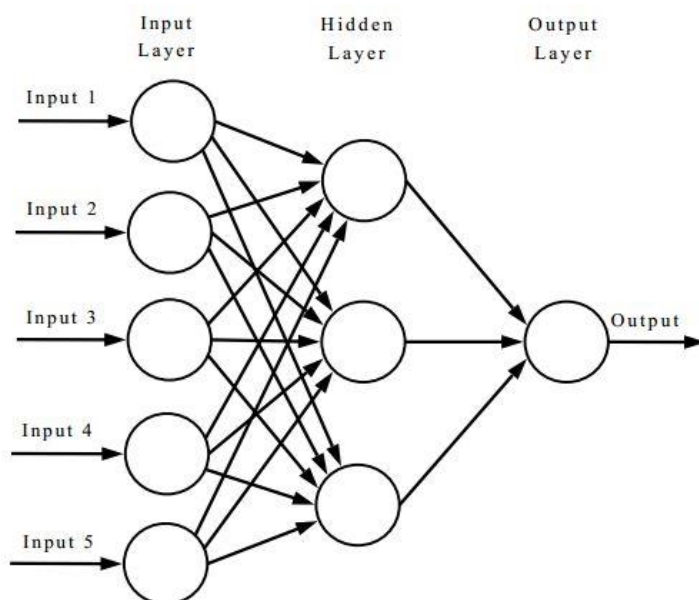


图 3-3 神经网络模型示意图

如图所示，输入层主要是接收输入的数据，在隐藏层存在激活函数将其进行计算然后再进行输出工作。

假设本文有一个由 M 个输入变量构成的向量 $X \in \mathbb{R}^{1 \times M}$ ，神经网络有 N 个的神经元，只有一个最终的输出变量 $y \in \mathbb{R}^{1 \times 1}$ 。那么在输入层进入隐藏层的时候，神经网络是进行下述计算的：

$$z = xw^T + b_0, \quad w \in \mathbb{R}^{N \times M}, b_0 \in \mathbb{R}^{1 \times N}$$

每个神经元都有一个常数变量 b_0 ，在隐藏层接收到 z 后就开始运行激活函数，常用的激活函数有 **sigmoid function**, **relu function**。需要注意的是，激活函数都是选用非线性的函数，因为线性函数关系无论叠加多少层，对解决问题都是没有意义的，同时运用神经网络解决的问题常常都是非线性问题（例如分类问题）。总的来说，神经网络的计算总公式就为：

$$y = \sigma(xw^T + b_0)\theta + b_1$$

其中，

$$w \in \mathbb{R}^{N \times M}, b_1 \in \mathbb{R}^{1 \times N}, \theta \in \mathbb{R}^{N \times 1}, b_1 \in \mathbb{R}^{1 \times 1}。$$

卷积神经网络(**Convolutional Neural Networks**, **CNN**)是一种前馈神经网络，由若干个卷积层、池化层以及全连接层组成。卷积的结构使得 **CNN** 能够很好地利用输入数据的二维结构处理图像和语音数据。卷积层的功能是对入的音频特征参数进行进一步特征提取，其内部包含多个卷积核，然后通过池化层进行特征选择和信息过滤，输入到全连接层解除多维结构展开为向量，并通过激活函数传递到下一层网络。经过最后一个全连接层后，使用 **softmax** 函数输出分类结果。

卷积神经网络中，值得一提的是池化层。池化层具有特征选择与过滤的功能，主要目的就是降低特征图的特征空间。常见的池化操作为最大值池化、平均值池化和 **L2** 池化。但是本文因为是使用的双通道音频数据，故而选择了 **L1** 池化（可以理解为对序列数据的池化操作），卷积神经网络的一种结构如下图，但是这个结构是可变的：

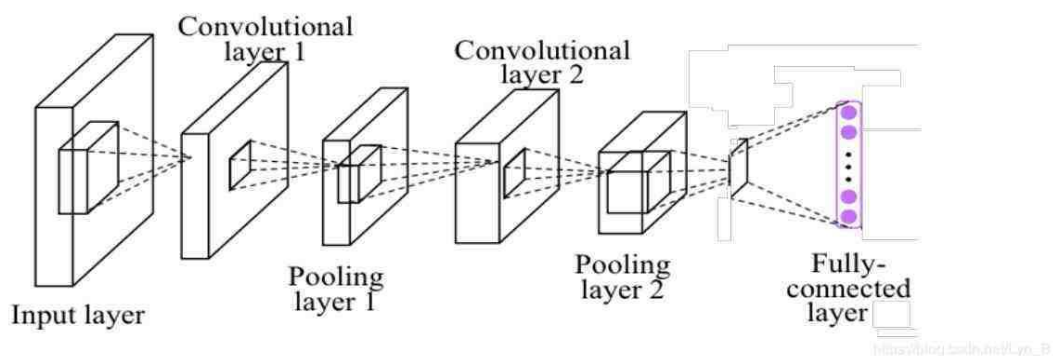


图 3-4 卷积神经网络示意图

音频数据在计算机中的存储是一个极长的矩阵序列，以一首 4min 的歌曲为例，现行歌曲大部分都为二声道歌曲，其在计算机中的存储为一个 $2 \times 20000000+$ 的一个超长矩阵，类似于图像的计算机存储（通道数 \times 矩阵维度），但是却又比图像存储低一个维度。

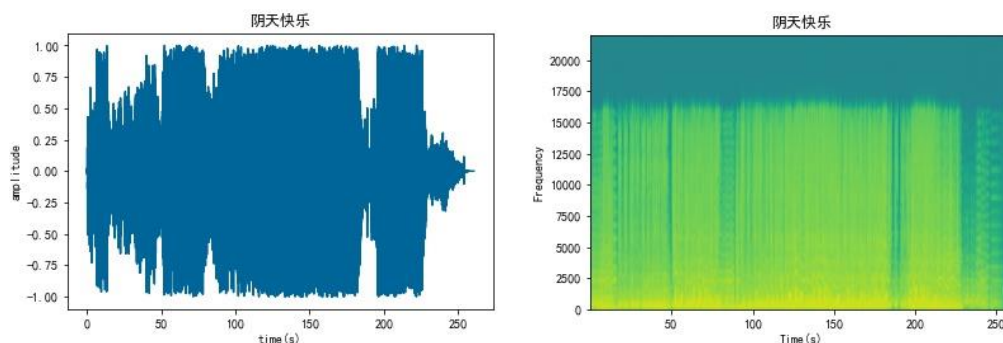


图 3-5 音频数据可视化示意图

本文为了提升推荐效果，提升模型科学性，决定采用卷积神经网络算法。

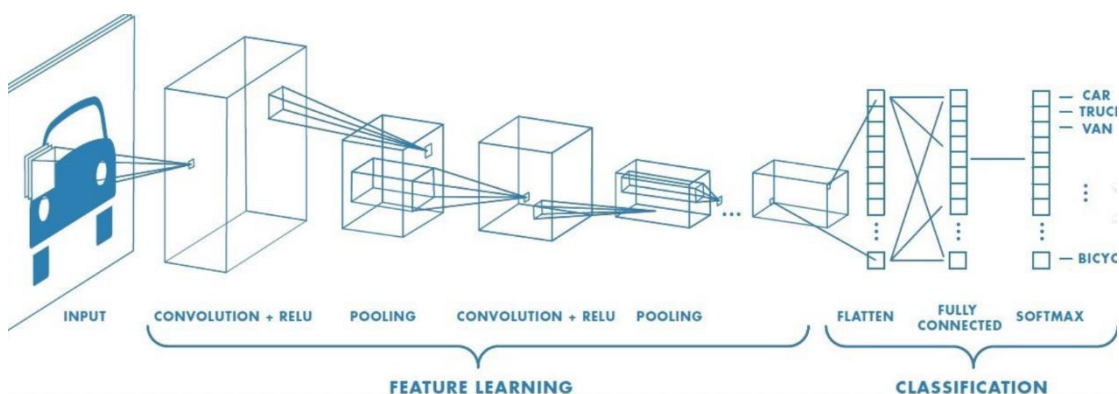


图 3-6 视觉领域常用网络

如图为视觉领域常用的卷积神经网络，本文在构造本文的音频网络时，对视觉网络进行了若干改动，除了全连接层和 Dropout 层以外，本文将卷积层进行降维以适应音频数据，并且对池化层进行了适应音频数据的改动，同时应用 Reshape 层进行数据维度调整，最终建立了本文的卷积神经网络模型。

通过本文的构建卷积神经网络模型，本文构建并且得出了用户大的测试结果，通过划分测试集对模型效果进行评估，发现准确率为 0.69。

3.3 融合多态信息的卷积神经网络推荐算法

模态是指事物发生或存在的方式，多模态是指两个或者两个以上的模态的各种形式的组合。目前研究领域中主要是对图像，文本，语音三种模态的处理。概括来说，多模态的显著特点是：冗余性和互补性，不同模态的表现方式不一样，看待事物的角度也不同，因此存在一些交叉（即信息冗余），且不同模态对事物的描述有

互补部分，所以比单个特征更优秀，甚至模态间可能还存在多种不同的信息交互，如果能合理的处理多模态信息，就能够得到丰富的特征信息。多模态学习可以划分为五个研究方向：多模态表示学习、模态转化、对齐、多模态融合、协同学习，本文在本文中主要使用了多模态融合。

为了让模型更加完全，在本文中，本文基于多模态构建了本文的推荐模型，融合了文本特征、音律特征以及图论特征，将三种模态的特征映射到一个向量之中。多模态特征的融合文本和图像的方法主要有三种：基于简单操作的，基于注意力的，基于张量的方法。本文中本文使用了简单操作融合办法。

多模态机器学习，英文全称 **MultiModal Machine Learning (MMML)**，旨在通过机器学习的方法实现处理和理解多源模态信息的能力。本文在此引入了文本信息（歌词），通过 **Python** 爬虫获取了大量歌曲的歌词，在将歌词切词后进行情绪分析，从而把文本特征与听觉特征结合在一起进行辅助分析。

通过多模态，本文重新构建了卷积神经网络算法，并且进行了模型预测，发现此时的准确率为 **0.75**，说明多模态有显著提升了模型效果的能力，说明歌曲的歌词也是影响用户歌曲偏好的重要决策因素之一。

3.4 基于支持向量机的歌曲推荐

支持向量机 (**Support Vector Machines, SVM**)，使用铰链损失函数 (**hinge loss**) 计算经验风险 (**empirical risk**) 并在求解系统中加入了正则化项以优化结构风险 (**structural risk**)，是一个具有稀疏性和稳健性的分类器并可以通过核方法 (**kernel method**) 进行非线性分类，是常见的核学习 (**kernel learning**) 方法之一。其优点包括：对复杂的非线性边界的建模能力好；不太容易过拟合；可以很好地应用于高维数据；适用范围广泛等

原理：最大间隔分类器：找到一个超平面，使得所有的实心圆位于超平面一侧，而空心圆位于另一侧，可能存在无穷多个这样的超平面。**SVM** 的目的就是想要找到一个具有最大分类间隔的超平面，以“最好地”区分这两类点，以至如果以后有了新的点，这条线也能做出很好的分类。

SVM 主要包括线性和非线性两类，本文中，本文主要运用非线性 **SVM** 算法进行判别模型的构建。非线性函数将输入数据映射至高维空间后应用线性 **SVM** 可得到非线性 **SVM**，非线性 **SVM** 中，核函数的选择是影响 **SVM** 最大的变量。常用的

核函数有线性核、多项式核、高斯核、拉普拉斯核、sigmoid 核，或者是这些核函数的组合。这些函数的区别在于映射方式的不同。

显然，在进行歌曲推荐时是一个二分类问题，具体过程如下：

- 多模态融合特征，即将歌曲的音频特征和文本特征融合。
- 标记类别，将出现在用户歌单中的歌曲的类别标为 1，未出现在用户歌单中的歌曲的类别标为 0。
- 划分训练集、验证集、测试集，进行 SVM 模型的训练和预测，其中测试集歌曲类别均为 0。
- 从测试集中寻找预测类别结果为 1 的歌曲作为推荐。

3.5 关于数据不平衡的补全方法——DCGAN

生成对抗网络是一种深度学习模型，是由蒙特利尔大学的 Ian Goodfellow 在 2014 年提出的机器学习架构。作为一种无监督学习的算法，生成对抗网络主要分为生成模型和判别模型。在 Ian Goodfellow 的论文中，他将生成模型类比为团队当中的伪装者，而判别模型就担任警察的角色，负责找出其中的伪装者。当判别模型找到伪装者的概率较小时，就说明生成模型生成的数据更接近真实数据，就达到了生成对抗网络模型的目的。其目标函数为：

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))]$$

$D(x)$ 代表的是 x 来自真实数据而不是生成器生成的数据集的概率， G 是带有参数的多层感知机表示的可微函数。

所以对抗生成模型的主要思路就是生成模型生成数据后，然后通过判别模型对生成模型进行判断，接着生成模型根据旧的判别模型的结果进行改进获得新的数据，判别模型再次进行学习进行判别工作，通过多次以上的博弈场景，最后达到的效果为：生成模型和判别模型都无法再改进，就生成了一个比较满意的模型。

DCGAN 与 GAN 不同的地方在于 DCGAN 将 G 的多层感知机模型改为了卷积神经网络，同时，为了使网络可微，拿掉了 CNN 中的池化层，另外将全连接层以全局池化层替代以减轻计算量。总之，DCGAN 成为一个生成模型和判别模型都运用了深度卷积神经网络的生成对抗模型。

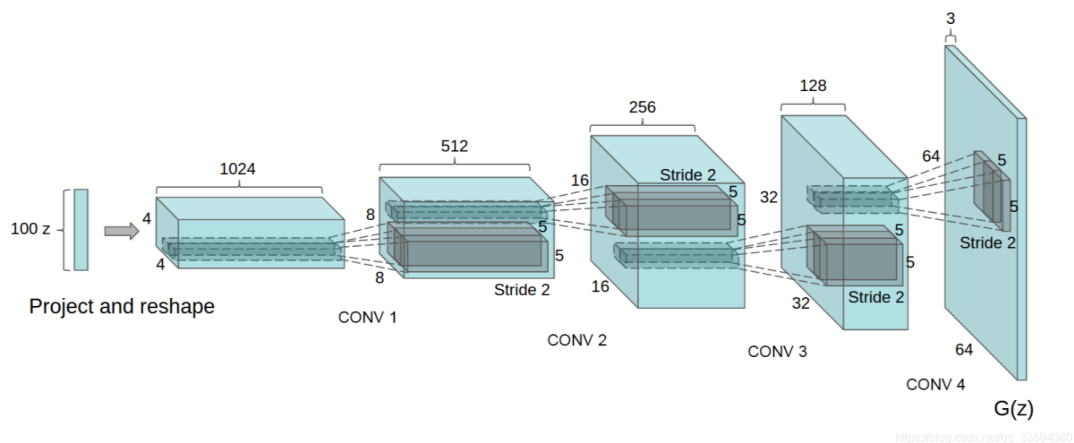


图 3-7DCGAN 示意图

DCGAN 相比于 GAN 或者是普通 CNN 的改进包含以下几个方面：

- 使用卷积和去卷积代替池化层
- 在生成器和判别器中都添加了批量归一化操作
- 去掉了全连接层，使用全局池化层替代
- 在生成模型和判别模型中使用 **batchnorm**。

考虑到个人机器算力约束，本文在此以某用户喜欢的歌曲为例，训练生成出来了一段音乐，作为数据集的补充，降低数据不平衡带来的问题。

3.6 协同聚类

基于歌曲的协同过滤主要考虑歌曲之间的相似性，通过计算推荐池中的歌曲与用户歌单中歌曲的音频特征的相似性，并对其进行排序，从而选取相似性最高的歌曲作为推荐。

由于用户数量较多，并且每个用户的歌单中又包含较多歌曲，若直接计算每首歌与推荐池中所有歌曲的相似性，需要运行较长时间，因此本文先将推荐池中的歌曲进行聚类，从而对于用户歌单中的歌曲，先在推荐池中寻找相似性最大的类别，再从该类别中寻找相似性最大的歌曲作为推荐。

实际上，对于用户歌单中的歌曲，并不是所有歌都能够从推荐池中找到与之相似的歌曲，因为本文只对推荐池中的歌曲进行了音频特征的提取，而用户歌单中的歌曲有可能不存在于推荐池中——即无法获取其音频特征来计算相似性。所以，最

终的基于歌曲聚类的协同过滤推荐机制如下：

将推荐池中的歌曲分成 k 类；

- 设用户歌单中共有 n 首歌曲，令 $i=0$ ；
- $i=i+1$ ，若 $i>n$ ，则结束推荐，否则判断歌曲 i 是否存在于推荐池中，若存在，则进入 Step4，否则，重复 Step3；
- 找到歌曲 i 在推荐池中所属的类别，计算歌曲 i 与该类别中其余歌曲的相似性，并进行排序得到序列 M ；
- 若序列 M 的长度等于 0，则进入到 Step6，否则从序列 M 中选择与歌曲 i 相似性最大的歌曲，若其还未被推荐并且不存在于用户的歌单中，则将其作为推荐，否则将其从序列 M 中剔除，并重复 Step5；
- 若 $i<n$ ，则回到 Step3，否则结束推荐。

4.总结

ZETDARD 音乐推荐系统是为了实现高质量的精准推荐，节省时间的基础上提高用户体验感和用户留存率；对于一些听歌较少的用户，基于生成对抗网络生成相关数据，扩大训练集。

此项目的实现分为三个阶段，首先数据获取阶段，通过网易云 API 直接按需进行数据的调取，例如用户、歌曲、歌词、评论、歌单等；并运用 python 进行网页爬虫，实现歌曲相关信息与歌曲 ID 的一一对应关系，通过动态 js 加密爬取得到用户 id 和昵称的对应关系；同时对获取到的图片、音频等进行处理，将单调的数据可视化，提高了数据的可读性和延展性。

其次使用推荐系统算法模型，通过 python 的 networkx 包构建图论网络模型，并且通过图论算法为用户初步推荐其可能会喜欢的歌曲、用户、歌手。DCGAN 生成模型和判别模型都运用了深度卷积神经网络的生成对抗网络，在生成模型生成数据后通过判别模型对生成模型进行判别；通过卷积神经网络 CNN 对输入的音频特征参数进行特征提取，进而进行特征选择和信息过滤，最终分类结果。基于多模

态构建了推荐模型，融合了文本特征、音律特征以及图论特征，将三种模态的特征映射到一个向量之中。还加入了支持向量机，用 SVM 算法进行判别模型的构建。考虑到歌曲之间的相似性，使用了基于歌曲聚类的协同过滤，寻找相似性最大的歌曲作为推荐。由于获取数据过多，因此使用了用户——歌曲——歌手网络构造，运用图论算法为用户进行初步的推荐。通过聚类算法得到推荐池，构造并划分训练集、预测集、验证集，采用卷积神经网络来构造音频推荐算法。通过多模态，重新构建了卷积神经网络算法，并且进行了模型预测。标记类别在用户歌单上的歌曲和不在用户歌单上的歌曲，通过 SVM 模型训练和预测，进行歌曲推荐。

但此推荐系统也有一些不足之处，部分算法速度较慢，需要算力过大，不适用于业界真正使用，由此推知业界的推荐模型不应过为复杂，或许是图论/聚类为主；可以通过深度学习度量音乐家音乐风格和听歌人偏好风格，此处或许可做聚类，讲歌手推荐和伴侣推荐深度学习化。

参考文献

- [1]陈长风.基于 CNN-LSTM 的歌曲音频情感分类[J].通信技术,2019,52(05):1114-1118.
- [2]龚志,邵曦.基于多模态的音乐推荐系统[J].南京信息工程大学学报(自然科学版),2019,11(01):68-76.
- [3]李春阳,李楠,冯涛,王朱贺,马靖凯.基于深度学习的洗衣机异常音检测[J].山东大学学报(工学版),2020,50(02):108-117.
- [4]付炜,杨洋.基于卷积神经网络和随机森林的音频分类方法[J].计算机应用,2018,38(S2):58-62.
- [5]Transactions of the International Society for Music Information Retrieval Volume 3, Issue 1. 2020.
- [6] 该 API 的项目地址:
[Binaryify/NeteaseCloudMusicApi](https://github.com/Binaryify/NeteaseCloudMusicApi): 网易云音乐 Node.js API service (github.com)

致谢

不知不觉，一晃四年转瞬即逝，学校，学院，老师，辅导员为我个人在平时的生活、学习和升学中给予了莫大的宽恕和帮助，我也通过这四年从一无所知的学生成长为一个有着充分技能的学士，谢谢老师，谢谢学院，谢谢学校！

