



闽江学院

本科毕业论文（设计）

题 目	基于大数据个性化音乐推荐
学 生 姓 名	黄馨云
学 号	3187103113
学 院	数学与数据科学学院（软件学院）
年 级	2018 级
专 业	软件工程（闽台合作）
指 导 教 师	曹永忠
职 称	助理教授
完 成 日 期	2022 年 5 月 8 日

闽江学院毕业论文（设计）诚信声明书

本人郑重声明：

兹提交的毕业论文（设计）《基于大数据个性化音乐推荐》，是本人在指导老师 曹永忠 的指导下独立研究、撰写的成果；论文（设计）未剽窃、抄袭他人的学术观点、思想和成果，未篡改研究数据，论文（设计）中所引用的文字、研究成果均已在论文（设计）中以明确的方式标明；在毕业论文（设计）工作过程中，本人恪守学术规范，遵守学校有关规定，依法享有和承担由此论文（设计）产生的权利和责任。

声明人（签名）：

2022年05月11日

摘 要

由于搜索引擎的快速发展，衍生了很多的推荐系统。近年来，音乐领域蓬勃发展。能在海量的音乐中快速有效的得到个人所喜欢的音乐资源，是一件很重要的事情。

网易云音乐针对用户的个性化需求，为用户提供了可以发表自己的音乐感受和评论的平台。使用者可以根据自己的喜好，将自己喜欢的歌曲分享给别人。

基于网易云音乐的海量数据资源，运用大数据技术进行数据分析，主要模块分为数据收集、数据存储、数据分析、数据展示四大模块。

个性化音乐推荐系统是一般推荐系统更加复杂的系统，结合网易云音乐的海量资源，探索用户行为背后的规律，同时挖掘音乐相关的评论话题，了解用户喜好以及音乐的风格类型，从而建立了基于大数据个性化音乐推荐的数据分析平台。

关键词：大数据分析；网易云音乐；音乐数据分析平台

Abstract

Due to the rapid development of search engines, many recommendation systems have been derived. In recent years, the music field has been booming. It is very important to be able to get your favorite music resources quickly and effectively in the mass of music.

NetEase Cloud Music provides a platform for users to express their music feelings and comments according to their personalized needs. Users can share their favorite songs with others according to their preferences.

Based on massive data resources of NetEase Cloud Music, big data technology is used for data analysis. The main modules are divided into four modules: data collection, data storage, data analysis and data display.

Personalized music recommendation system is more complex than general recommendation system system, combined with the enormous resources netease cloud music, explore the law behind the user behavior, mining the comment music related topics at the same time, the understanding of user preferences as well as the style of music types, which based on data analysis platform for large data personalized music recommendation.

Key words: Big data analysis; NetEase Cloud Music; Music data analysis platform

目 录

1 引言	
2 音乐互动平台	
2.1 音乐互动平台的概念	- 2 -
2.2 什么是音乐社交	- 2 -
2.3 音乐的发展现状	- 3 -
2.4 移动互联网市场现状	- 3 -
3 大数据分析平台	
3.1 大数据基本特征	- 4 -
3.2 大数据技术介绍	- 4 -
3.2.1 网络爬虫采集技术	- 4 -
3.2.2 MongoDB 分布式存储技术	- 5 -
3.2.3 Spark 分布式计算技术	- 6 -
3.2.4 Java Web 技术	- 7 -
3.2.5 数据可视化技术	- 7 -
3.3 系统设计	- 7 -
3.3.1 数据收集模块的设计	- 8 -
3.3.2 数据存储模块的设计	- 11 -
3.3.3 数据分析模块的设计	- 12 -
3.3.4 数据展示模块的设计	- 12 -
4 数据多维度分析	- 13 -
4.1 在线音乐行业市场规模及发展形势	- 13 -
4.2 在线音乐用户分析	- 13 -
4.3 用户活跃时段分析	- 14 -
4.4 歌手及歌曲分析	- 15 -
4.5 音乐评论话题分析	- 17 -
4.6 音乐推荐	- 17 -
4.7 中国在线音乐产业发展现状分析预测	- 18 -

5 大数据技术在音乐发展模式上的主要问题和未来展望

6 结论

7 未来展望

参考文献 - 22 -

附录 1 - 23 -

致谢 - 24 -

基于大数据个性化音乐推荐

黄馨云

(闽江学院 数学与数据科学学院 (软件学院), 福建 福州 350108)

1 引言

伴随着互联网时代的到来, 音乐出现了数字化的趋势, 由于信息技术飞速发展, 数字化音乐也迅速崛起^[1], 随着网络音乐的迅猛发展, 数字化音乐数量与用户数量也呈现了爆炸式的增长。网易云音乐的特色就是提供个性化的推荐服务, 这是一个以个人喜好为主的网站, 其中还包含了许多有音乐鉴赏能力的人。网易云音乐最具有特色的功能便是可以在网易云音乐平台中进行音乐社交和音乐互动。

在这个崇尚个性化的年代, 网易云音乐突破了流行的传统音乐, 推出了面向不同人群的音乐推荐。从而, 获得了大量的用户和音乐资源。

在海量数据的支撑下使得网易云的数据分析得以数据可视化, 通过对网易云音乐中用户行为的潜在规则以及其音乐风格的类型和偏好进行深入的研究, 对音乐的主题进行深入的挖掘, 从而建立了基于大数据分析的个性化音乐推荐的数据分析平台。对网易云音乐的社交化策略进行有针对性和建设性的提升, 并且展望其发展前景。有效地使用推荐系统是解决信息超载问题的一个有效途径, 这使人们能够从复杂的信息流中挖掘出与自己感兴趣的相关信息。所以, 将推荐系统和音乐相结合是一种必然的发展趋势^[2]。随着移动互联网和数字多媒体技术的迅速发展, 数字音乐已经成为了一种主流的消费方式, 而基于个性化的音乐推荐也逐渐成为了当前的热门话题^[3]。

主要流程是利用网络爬虫技术进行数据的收集, 然后将收集到的数据存储到分布式文档数据库 MongoDB 上, 再使用 Spark 分布式计算技术对数据进行分析, 最后将分析的数据存储到 MySQL 中, 并进行可视化的展示。

2 音乐互动平台

2.1 音乐互动平台的概念

音乐互动平台的主要功能是对音乐进行发现和分享等，用户在音乐搜索、收听音乐的同时，也可以通过音乐网站进行音乐交流。在我国的数字音乐时代的背景下，数字音乐行业的“社会化”、“移动化”和“云服务”是其三大特点。随着智能手机的普及，数字音乐在智能手机上的应用将会是我国广大网民获取音乐的最主要和最普遍的途径。

音乐互动平台更是让人们有了一个可以互相交流，可以互相分享音乐的一个新媒介。人们在音乐互动平台上，可以对音乐进行搜索、交流、分享、评论、点赞、收藏，也可以对歌手进行搜索、分享、评论、点赞、收藏等互动交流。音乐互动平台使热爱音乐的人有了更多机会去认识热爱相同音乐的人，去结交到更多志同道合、品味相似的好朋友。

2.2 什么是音乐社交

音乐能轻易把人连接起来，它是一种极易引发共情的介质；同时，音乐品味也是一种相当个性化的东西，千人千面。音乐和社交，近来不断被共同提及的两件事。在我国数字音乐时代的大背景下，智能手机所带来的社会属性和音乐应用软件的社会属性可以进行互动。作为一个群体，社会是人成为人的必要条件，也正是由于它，传播才能被创造出来^[4]。在数字音乐时代，人们可以用智能手机、因特网迅速地把自己喜欢的歌曲与别人共享。在数字音乐时代的背景下，空间和时间层面上的限制被人们所打破，实现了个体之间的相互交流，接收和传递基于个体利益的信息。在微信上，用户可以将自己喜爱的歌曲或者评论，进行链接或者截图，这样既能和好友们交流，又能在音乐上结识一些有兴趣的人。我们可以把“音乐社交”定义为：以用户的社会交往为纽带，以用户生产内容为核心，最大化地开发用户的创意，并最终帮助用户获得音乐的享受和消费^[5]。

网易云音乐抓住这个机会，创新推出了“音乐社群”这个概念，并从一个全新的视角，诠释了“音乐”的含义，以用户为基础，打造一个完整的音乐生态圈。网^[6]。

2.3 音乐的发展现状

伴随着移动互联网时代的到来，以及智能手机的日益普及和应用，越来越多的人开始选择下载、使用互联网音乐软件享受音乐^[7]。科技进步让人们在享受数字化服务带来的便利时，极大程度地促进了互联网的发展。在互联网的大环境下，越来越多的音乐应用软件出现在普通民众的生活之中，网络的普及应用，生活方式也发生了日新月异的变化，数字音乐这个产业，近年来蓬勃发展。在网络时代，随着网络的发展，以及智能手机的普及，数码音乐的发展也得到了极大的发展，这使得传统音乐迎来了更多的机遇和挑战。

随着手机的普及，人们对电子终端的依赖性越来越强，网易云音乐的开发模式也越来越多，以此迎合用户的需要。在这种情况下，在线音乐平台拥有了巨大的发展空间，大量的音乐软件进入了中国的网络音乐市场，这就导致了原本已经饱和的市场趋于“饱胀”。

网易云音乐的兴起，既是手机和移动互联网的发展，也是数码音乐行业发展到今天的必然结果。网易云音乐在已经饱和的音乐市场中走出了一条新的道路，开辟了属于自己的新的模式。

网易云音乐的社会化发展理念，推动了网易云音乐实现了以用户为本、服务为主的理念，让网易音乐在大多数音乐类型应用软件中独占鳌头。

2.4 移动互联网市场现状

移动互联网移动互联网是移动通信和传统互联网融合的产物^[8]，它是将互联网平台、技术、业务模式和应用与移动通讯技术相结合和实施的行为的统称。使用者通过无线通讯网路，利用移动电话或其他手持式终端机来上网。

移动互联网是网络终端向个体的延伸，是手机与互联网的结合，它正逐步渗透到人们的生活和工作中。各类应用软件和其他丰富多彩的应用程序近年来迅速发展，并且深刻地影响着信息时代人们的生活。

3 大数据分析平台

3.1 大数据基本特征

第一个特征是数据量大(**Volume**)。大数据的起始计量单位至少是 P(1000T)、E(100 万 T) 或 Z(10 亿 T)，数据的收集，运算，存储，都是一个巨大的数字。

第二个特征是数据类型繁多(**Variety**)，且资料的种类和来源都很丰富。主要有：结构化、半结构化、结构化、非结构化，常用的有：音频、图片、视频、网络日志、地理位置等等。

第三个特征是价值密度低(**Value**)，且具有相对低的数据价值密度。信息量很大，但是数据价值密度相对较低，就像是百炼成钢，从海量的信息中提取出了一些有价值的信息。

第四个特征是时效高、速度快、实时、快速(**Velocity**)，时效性要求高，处理速度快。数据的快速增长，快速的处理，以及快速的获得。这就是数据挖掘和大数据最明显的不同之处。

3.2 大数据技术介绍

3.2.1 网络爬虫采集技术

网络爬虫，也叫 **web 蜘蛛**，也叫 **web 机器人**，是一种可以根据特定的规则，通过某种特定的方式，来抓取万维网中的数据程序。

网络爬虫是通过因特网上的服务器来爬取获取内容。这是一种用电脑语言写成的程序或脚本，可以从因特网上自动取得资讯或资料，在每一页上搜寻和捕捉特定资讯，直至完成所有可以正常开启的网页为止。

网爬技术已经成为这个时代先进互联网技术的一部分，用户需要通过数据来分析用户特征、产品的不足之处以及对手信息等，数据的采集是爬虫技术重要的基础部分。

大数据盛行的当下，该技术是一个不可缺少的环节，人们需要数据去对用户行为、产品的缺陷、竞争者的信息来进行分析，而这些都是由数据收集来完成的。

所谓的网络爬虫，其实就是“数据”，在这个互联网时代，信息是无价之宝，什么都是数据，谁掌握了海量的信息，谁就能掌握更多的信息。在搜索引擎、数据采集、广告过滤、大数据分析等方面，网络爬虫的应用非常广泛。

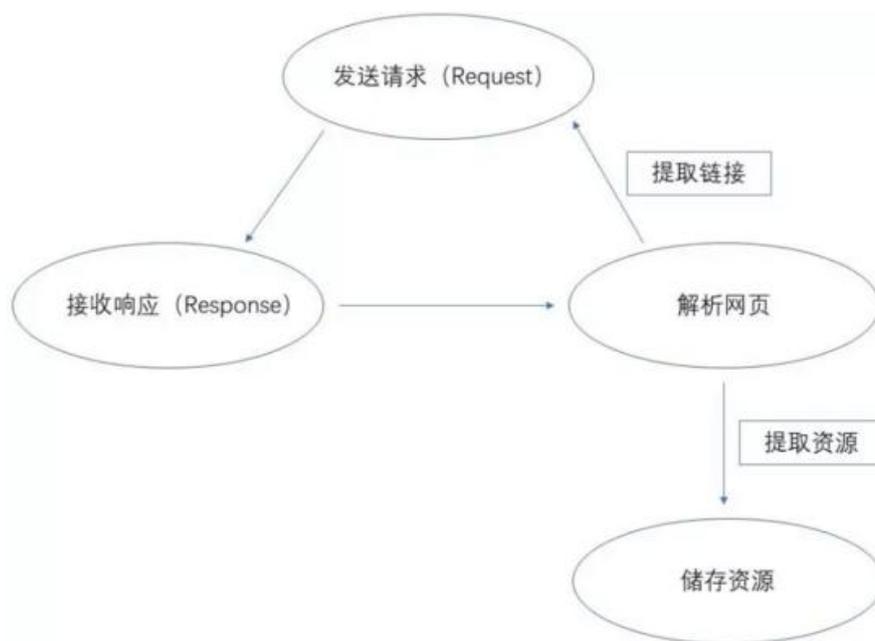


图 3-1 爬虫的基本原理

3.2.2 MongoDB 分布式存储技术

MongoDB 是一个基于分布式，文件存储的非关系型数据库，其特点是高性能、易部署、易使用，存储数据方便^[9]。其目的是为 WEB 应用程序提供可伸缩、高性能的资料储存解决方案。

MongoDB 是一种介于关系型和非关联型型的非关系型型数据库。它所支持的数据结构更宽松，与 json 的 bson 格式相似，所以能够存储更繁杂的数据类型。MongoDB 最大的特色就是支持更多的查询语言，其语法与面向对象的查询语言有些相似，能够完成与关系数据库相似的单一表格查询，并且能够提供数据的索引。

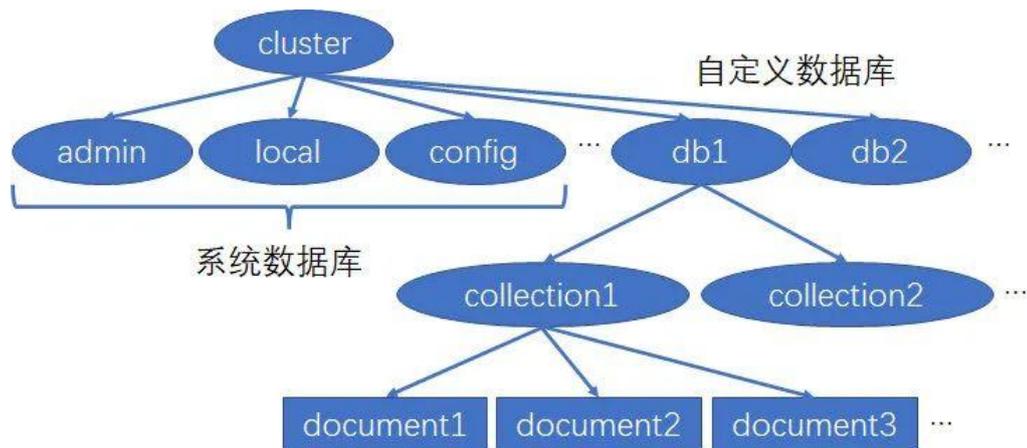


图 3-2 MongoDB 实例的数据结构

3.2.3 Spark 分布式计算技术

推荐系统要对大量的数据进行筛选，并且要处理大量的数据，因此必须具有很强的大数据处理能力。而其中 Spark 是最新一代的计算框架，大数据处理能力很强，使用 Spark 将大大提高推荐系统的运行效率^[10]。

Spark 不但具有高效率的内存操作能力，同时也是一套通用的软件堆叠，它可以应用于各种应用场合，这些应用程序本来就要求使用许多不同的分布式平台。Spark 是一种用于大规模数据运算的通用分布式大数据处理引擎，它是一种应用于大数据处理的基础技术。Spark 相对 Hadoop 更有优势，但是 Spark 无法完全替代 Hadoop。实际上，Spark 作为 Hadoop 的一个成员，在 Hadoop 中，它的作用是取代 MapReduce 计算模型。

Spark 具有 Hadoop MapReduce 所有的优势，但是区别在于，每一次 Hadoop 通过 job 的中间处理都会被存储在 HDFS 上，而 Spark 执行 job 的中间处理数据则可以被直接存储在内存中，而不需要在 HDFS 磁盘上读取。因为内存的读写速度与磁盘的读写速度不在一个数量级上，所以 Spark 可以更快地进行数据的运算处理。

此外，由于 Spark 在内部使用了弹性分布式数据集(Resilient Distributed Dataset, RDD)^[11]，RDD 是 Spark 中的核心概念本质是一个泛型的数据对象，可以理解为数据容器，本身是一个复合型的数据结构。经过了数据模型的优化，因此即使是在磁盘上进行分布式运算，它的计算效率也要比 Hadoop MapReduce 高。

3.2.4 Java Web 技术

Java 已经发展了 20 多年，它是当前最受欢迎的一种编程语言，它可以为不同的用例和场景提供不同的框架，目的在于减少开发人员的代码和基础工作。重点介绍 Spring 框架在 Java Web 技术堆叠中的应用。Spring 基本上是通过 JavaBean 实现的，而这个 JavaBean 本来只能在 EJB 中实现。这是一个综合了 IOC 控制反转和面向切面 AOP 编程的容器框架，与此同时，Spring 框架是一种轻量级的 JavaEE 框架^[12]。Spring 不只是一种 Web 开发框架，而是基于 Java 技术堆叠的一套完整的编程模式，从 Spring Boot 起步，Spring 是一种无需应用服务器就能以最低配置启动和运行 Spring 应用的一种方法。

3.2.5 数据可视化技术

数据可视化技术是一种能够更直观地显示数据分析结果的大数据技术，一般采用折线图、词云图、柱状图、饼状图等方式来表示。通过数据可视化技术，人们可以更加清晰和直观地感受到数据的分析结果。

本文介绍所使用到的 Echarts，是百度开发的一个 JavaScript 图形库，该图形库可在 PC 机和手机上使用，与大多数浏览器兼容，提供丰富、直观、高度个性化的数据可视化图形。

3.3 系统设计

通过收集数据、存储数据、分析数据及展示数据，实现了基于大数据分析的个性化音乐推荐的数据分析平台。主要内容包括如下几个层面：

- 1) 通过网络爬虫对数据进行收集。
- 2) 对采集到的数据进行预处理，并将其存入数据库。
- 3) 对存储的数据进行多维分析。
- 4) 将分析后的数据进行可视化展示。

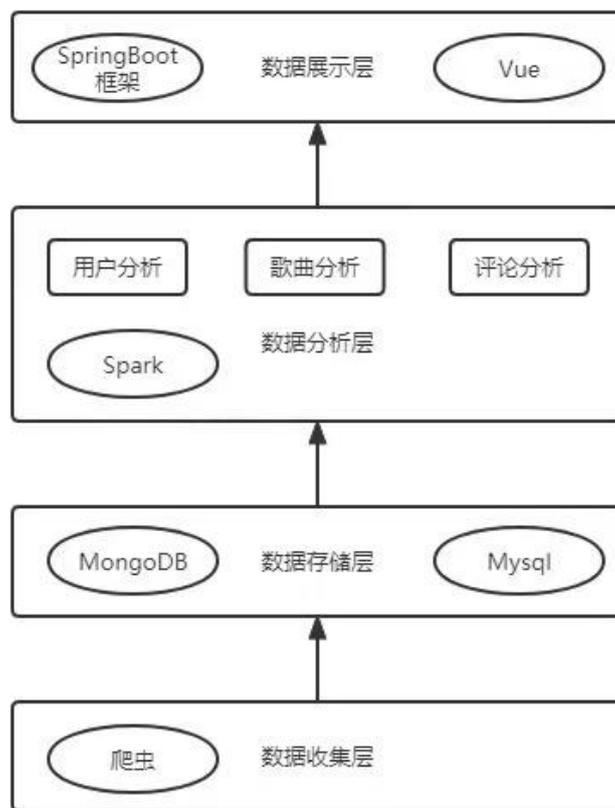


图 3-3 系统架构图

3.3.1 数据收集模块的设计

数据收集是非常重要的部分，缺少了数据收集这一部分，将无法进行之后的数据存储、数据分析、数据可视化。只有对数据进行准确地收集后，才可以进行之后的数据存储、数据分析以及数据的可视化。

数据收集部分，使用的是 Python 爬虫，对网易云音乐发起请求爬虫对应的数据，请求的发起是使用 http 库向目标站点发起请求，即发送一个 Request，并且通过 json 和 xpath 对数据进行解析，得到 log 文件，获取网易云音乐的数据内容。

爬取的数据内容主要是评论数据、歌手信息、歌曲信息这三部分。用户信息有用户名、用户 ID、用户年龄、用户性别、用户所在地区、个人介绍，评论信息有评论内容、评论 ID、评论点赞数、评论时间，歌手信息有歌手 ID、歌手名、歌手分类，歌曲信息有歌曲 ID、歌曲名称、歌曲 URL，所爬取的数据就包括这些内容。

表 3-1 爬取数据部分内容

```

评论数据: {
    "comment": "我想开一家酒吧, 店里深夜开店, 半夜打烊, 没有激荡的音乐, 没有水果茶奶茶, 只有民谣和酒, 喝下的是酒, 吐出来的是故事, 我想和每个客人分享年少的事, 路过店门口的人都会停住脚步, 只因为他们听到了酒般的民谣店名就叫 流浪客.今年 19 我想在 25 岁, 开在长沙.",
    "comment_id": "5460300336",
    "datetime": "2021-11-16 07:17:03",
    "praise": "4",
    "song_id": "1890467746",
    "user_age": "29",
    "user_city": "321100",
    "user_gender": "1",
    "user_id": "2006181477",
    "user_introduce": "网易云音乐人, 作词, 作曲",
    "user_name": "抖音里的乔馆"
},
{
    "comment": "建议听一下 superlove bachbeats",
    "comment_id": "5460225417",
    "datetime": "2021-11-16 00:27:03",
    "praise": "8",
    "song_id": "1896177952",
    "user_age": "14",
    "user_city": "310101",
    "user_gender": "1",
    "user_id": "1603002154",
    "user_introduce": "",
    "user_name": "欣依然在这"
},
{

```

```
"comment": "有感觉了噻",
"comment_id": "5460176327",
"datetime": "2021-11-15 23:27:03",
"praise": "1",
"song_id": "1895560037",
"user_age": "13",
"user_city": "350700",
"user_gender": "2",
"user_id": "3957192067",
"user_introduce": "我很帅 我知道",
"user_name": "云南白不药"
},
{
"comment": "也不是抄袭，就是直接拿来用了……",
"comment_id": "5460095869",
"datetime": "2021-11-15 21:54:03",
"praise": "1",
"song_id": "1896177952",
"user_age": "26",
"user_city": "320400",
"user_gender": "2",
"user_id": "1689619073",
"user_introduce": "",
"user_name": "wink 温客行"
},
```

歌手信息：

```
{"id": "1198010", "artist": "暗杠", "title": "华语男歌手 "}
{"id": "12127564", "artist": "艾福杰尼", "title": "华语男歌手 "}
{"id": "33047183", "artist": "阿吉野马", "title": "华语男歌手 "}
{"id": "30709773", "artist": "Abdulla Abdurehim 阿不都拉", "title": "华语男歌手 "}
{"id": "47450010", "artist": "艾热西丁", "title": "华语男歌手 "}
```

```
{ "id": "1876", "artist": "阿杜", "title": "华语男歌手 " }  
歌曲信息:  
{ "name": "小雨天气", "song_id": "id=1889702364",  
  "url": "https://music.163.com/#/song?id=1889702364" }  
{ "name": "漠河舞厅·2022", "song_id": "id=1894094482",  
  "url": "https://music.163.com/#/song?id=1894094482" }  
{ "name": "还是劈开", "song_id": "id=1861479524",  
  "url": "https://music.163.com/#/song?id=1861479524" }  
{ "name": "初恋", "song_id": "id=1873049720",  
  "url": "https://music.163.com/#/song?id=1873049720" }  
{ "name": "说几遍", "song_id": "id=1897113282",  
  "url": "https://music.163.com/#/song?id=1897113282" }  
{ "name": "小酒馆", "song_id": "id=1897083755",  
  "url": "https://music.163.com/#/song?id=1897083755" }
```

3.3.2 数据存储模块的设计

当爬取到数据，对数据进行收集后，需要对数据进行存储，方便后续对数据的使用。数据存储分为两个模块，分别是存储在 MongoDB 上和存储在 MySQL 中。爬虫所爬取的数据将存储在 MongoDB 上，以便后续数据分析使用。MySQL 中存储的数据主要是用于后续数据可视化展示使用。

MongoDB 分布式存储数据库功能强大，没有复杂的连接，使用方便。MongoDB 中保存的数据是评论数据、歌手信息数据、歌曲信息数据。之后数据分析，将使用 Spark 对 MongoDB 中的数据进行处理和分析。

MySQL 数据库中有多张表，包括年龄分析表、评论数据表、歌手信息数据表、歌曲信息数据表、用户信息数据表等。MySQL 中的数据将用于数据可视化展示，展示到 web 页面当中。



图 3-4 MongoDB 数据库表关系

3.3.3 数据分析模块的设计

当爬取到数据，并将爬取到的数据存储到数据库后，需要读取数据库中所存储的数据，并对这些数据进行处理分析。数据分析将使用 Spark 对评论信息数据、歌手信息数据、歌曲信息数据、用户信息数据进行处理及分析。

Spark 是一个以速度，易用性和复杂的分析为基础的数据处理框架，它为大量数据处理的需要（文本数据、图表数据等）和数据来源（大量数据或实时数据）。

数据分析主要包括用户在全地域分布情况、用户评论的活跃情况分布（日期中是否有假期等）、用户评论时间（每天小时分布）、用户情感分析、用户性别及年龄占比、歌手占比分布、歌手占比分布、热歌分布点赞情况。其中将使用 jieba 框架，对评论数据进行分词，得到用户使用频率最高的词语，使用停用词去除关系性不强的词语。

3.3.4 数据展示模块的设计

在将数据库中存储的所爬取到的数据进行处理和分析后，数据根据需求进行可视化展示，针对性的使用不同图表对不同的数据分析结果进行展示。

将分析后的数据存储到 MySQL 中，后端使用 Java 的 Spring Boot 框架，对 MySQL 数据库中的数据表进行读取。前端使用 Vue、Echarts 图形库来显示读入的资料，从而使人们更清晰、直观地了解资料分析的结果。

4 数据多维度分析

4.1 在线音乐行业市场规模及发展形势

互联网音乐在经历了数年的清理和整顿之后，从 2015 年开始，我国的互联网音乐进入了“正版时代”，并迅速走向了健康、商业化的道路。在中国，网络音乐已经成为一种主要的听歌方式，因为世界各地的录音音乐都在不断地减少。5G 时代的来临，在线音乐的流式播放方式将会出现音质更好、更便捷的发展时期，中国的网络音乐也将步入快速发展的轨道。从技术上来说，随着大数据等技术的普及，以及 5G 时代的来临，网络在线音乐平台的发展速度已经达到了一个新的高度，并推出了许多新功能比如个性化推荐^[13]。

在线音乐消费的主要趋势是内容的需求和支付意愿的改变。第一个方面是消费者的消费内容和消费场景的丰富。不过，现在的使用者已经不满足于单纯的音乐搜索和歌曲库，他们更多的是通过网络来发现好的歌曲，了解歌曲的资讯，找到有相同爱好的朋友，并抒发情感。第二个层面是消费者的消费渠道和消费理念的普及。随着移动电话越来越流行，网络越来越差，付款越来越方便，以及技术的进步，人们的收听方式基本上都是在线的。

4.2 在线音乐用户分析

根据数据分析，由图可知，随着互联网音乐用户的日益年轻化，年轻人更倾向于通过流媒体来满足自己的娱乐需要。根据网易云音乐用户画像，超七成的用户在 24 岁以下，有 46%的人在 19~24 岁，大部分大学生年龄在这个区间内^[14]。网易云音乐的主要使用者为 90-00 后，主要以 15-35 岁为主要人群。网易云音乐用户男女分布相对均衡，男女比例约为 4: 5。

用户覆盖面积广泛,各地区用户均有触及在全国范围内。用户主要分布在一二线城市,目前正逐步囊括三、四线城市用户。北京、广东、江苏三个省份的用户数量占全体用户的较大部分。

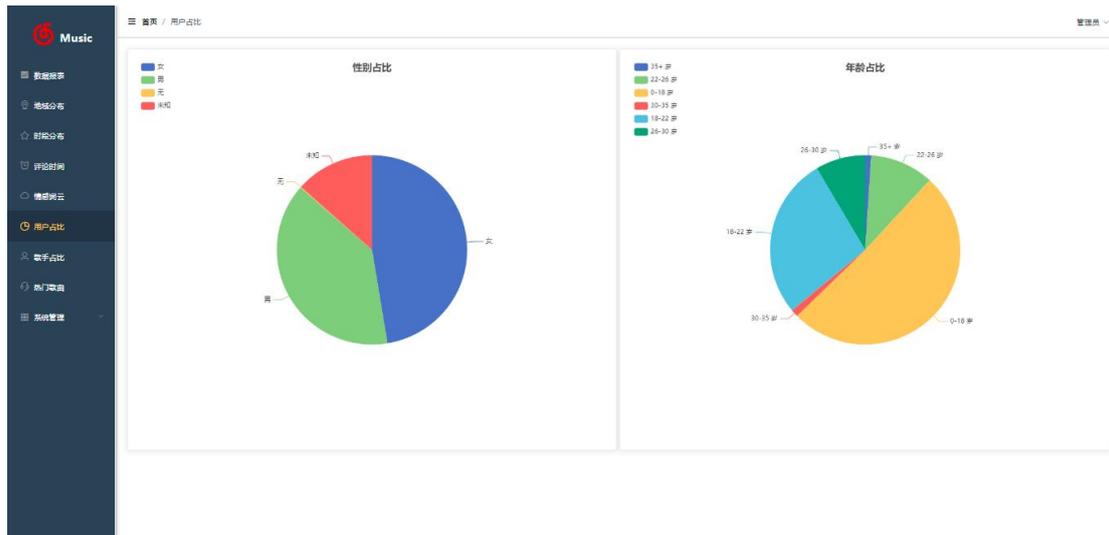


图 4-1 用户占比结果展示

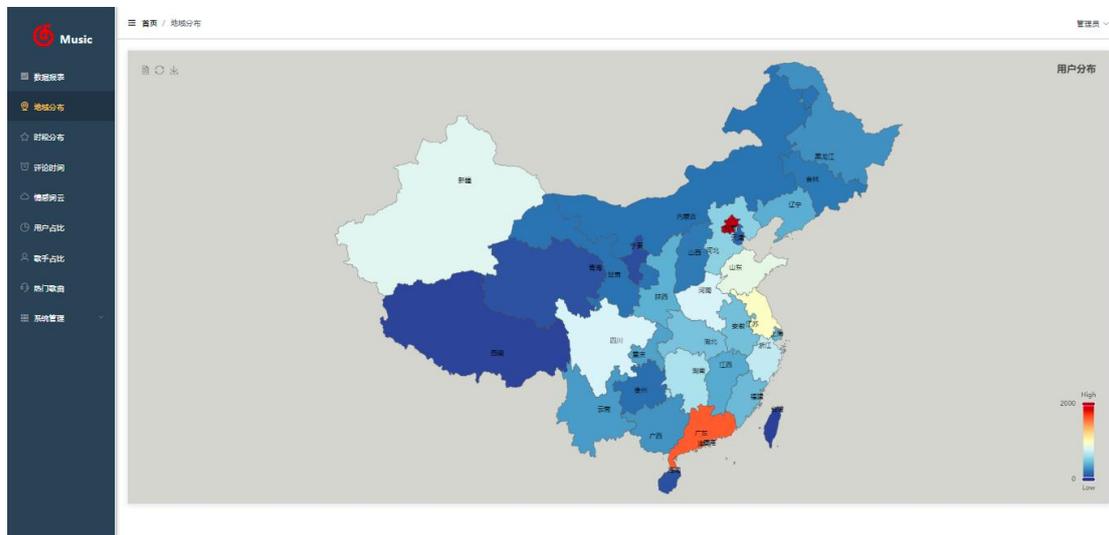


图 4-2 地域分布结果展示

4.3 用户活跃时段分析

由图可知，用户活跃时段主要集中于夜晚时段，网易云音乐用户年轻化特征明显，更爱晚睡晚起，其社交需求也相对更高，活跃时段主要集中在 20 时到 24 时区间，网易云音乐在 24 时后才出现回落，说明网易云音乐用户熬夜人群较多。网易云音乐用户主要集中在晚上进行活跃，在深夜，人们更容易情绪波动，产生一些比较伤感的情绪，听一些音乐歌曲。

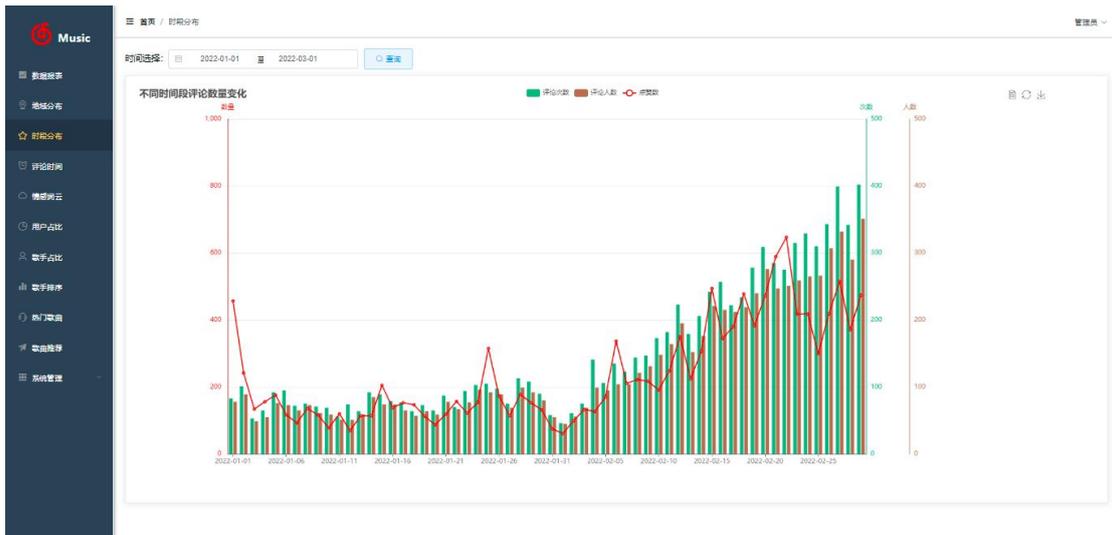


图 4-3 时段分布结果展示

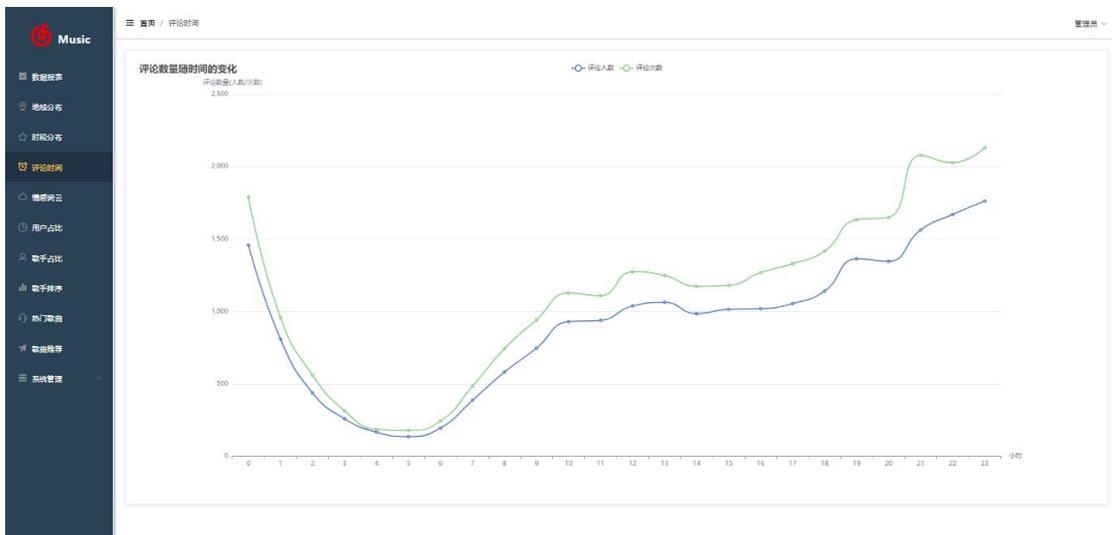


图 4-4 评论时间结果展示

4.4 歌手及歌曲分析

在线音乐的消费倾向是多样化的，除了主流音乐以外，用户也表现出了多样化的特征。华语流行曾经是网上流行音乐的主流形式，但是现在，随着年轻人的不断增长，各种类型的音乐都有很大的受众群，人们不再限于华语流行，日韩歌曲及欧美歌曲也深受年轻用户的喜爱,在网易云音乐的收听风格中排行前列。

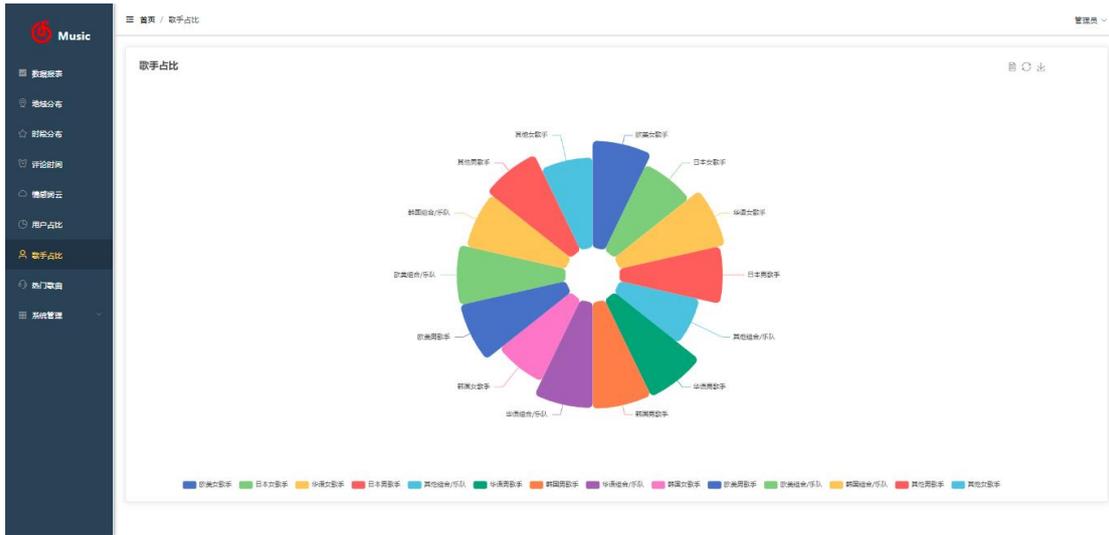


图 4-5 歌手占比结果展示

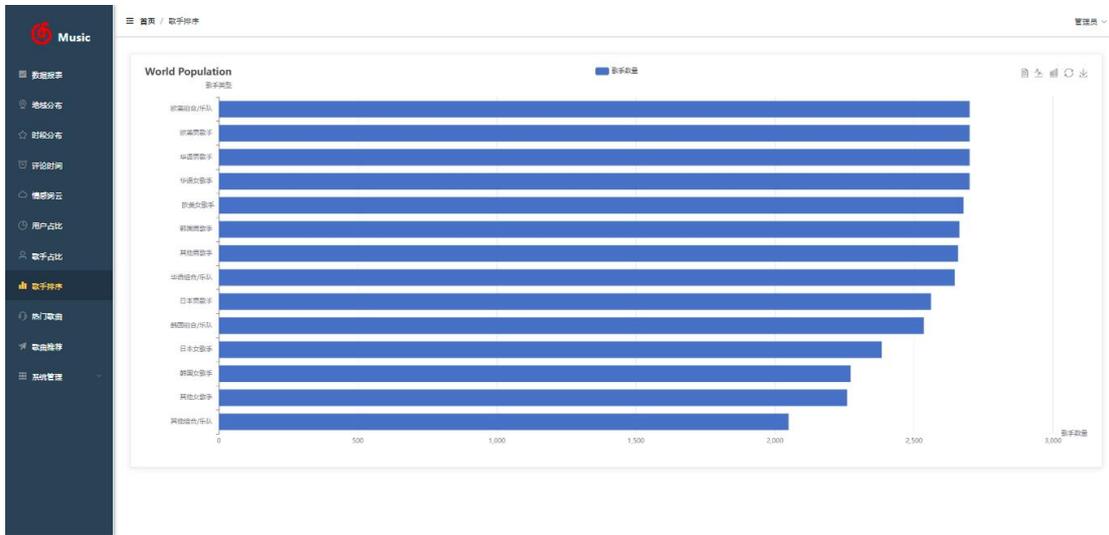


图 4-6 歌手排序结果展示

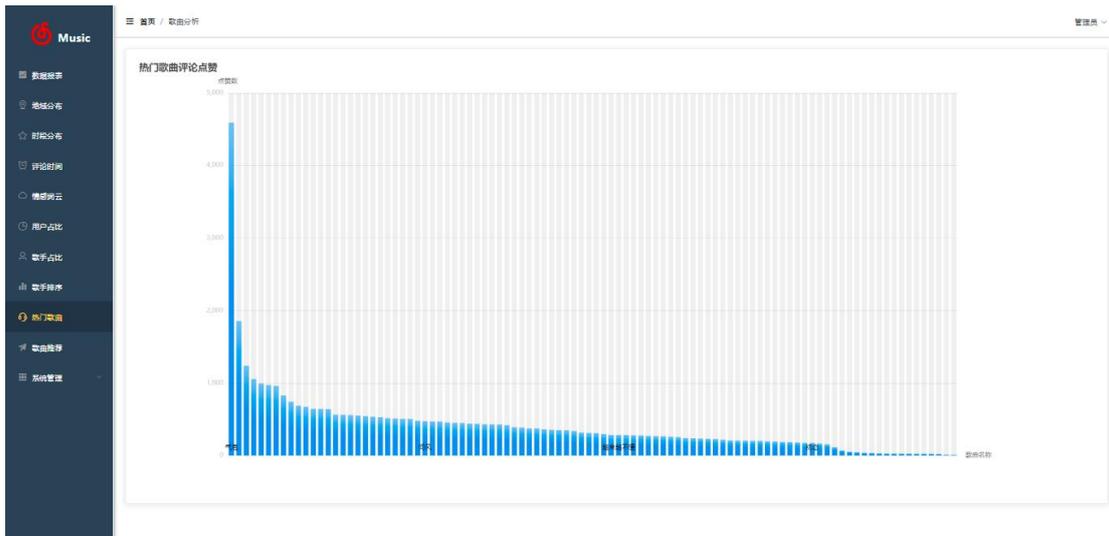


图 4-7 热门歌曲结果展示

4.5 音乐评论话题分析

音乐评论话题挖掘使用的是 Echarts 插件中的词云图，在词云图中，词语所占的面积越大，及代表这个词语的比重越大，所出现的次数越多。由图 4-8 可知，所占面积较大，使用频率较高的词语分别是好听、喜欢、首歌、多多、加油等词语。通过对用户的评价，以流行音乐为基础，试图为广大创作者的创作提供参考，以满足大众的口味^[15]。



图 4-8 评论分析结果展示

4.6 音乐推荐

用户可以通过歌曲推荐，根据歌手类型，歌曲类型进行搜索推荐，页面将根据用户的喜好，显示推荐符合条件的歌曲。其中，歌手类型有华语男女歌手、组合；欧美男女歌手、组合；日韩男女歌手、组合等。歌曲类型有民谣、摇滚、爵士、流行等。用户可以根据自己的喜欢进行搜索。

歌曲名称	歌曲专辑	歌手名称
Flower	流行	韩国歌手
꽃 (Flower)	流行	韩国歌手
Jasmine	流行	韩国歌手
Jam & Butterfly	流行	韩国歌手
Counting Star	流行	韩国歌手
Xi	流行	韩国歌手
Rubin	流行	韩国歌手
Downtown Baby	流行	韩国歌手
우리의 봄	流行	韩国歌手
지나지	流行	韩国歌手
남자중이	流行	韩国歌手
무제목을	流行	韩国歌手
비밀이나 숨겨	流行	韩国歌手
instagram	流行	韩国歌手
꿈이온기 (Carabini)	流行	韩国歌手
if i die	流行	韩国歌手
여우노래	流行	韩国歌手
사랑	流行	韩国歌手
백인 남자	流行	韩国歌手
나에게 가는 길 끝에서 (나의 길)	流行	韩国歌手

图 4-9 歌曲推荐结果展示

4.7 中国在线音乐产业发展现状分析预测

随着网络技术在中国的广泛应用，新的音乐产业也在不断地出现，在新的经济时代，音乐版权、运营、音乐内容制作等多个方面形成了一条完整的产业链。科技的发展促进了音乐产业的数字化，新媒体的涌现。同时，由于受到新冠肺炎的影响，中国的音乐界在线下运营，如现场歌剧、现场演唱会等都很难开展，面临运营上的风险；线上音乐发展云 Live、同屏 K 歌等新的商业模式，使得在线音乐产业得到了发展的机会。

中国在线音乐正在进入一个新的发展阶段，产品不断创新，版权逐步规范。科技的发展，内容的爆炸，使得用户在网络上的时间和金钱都有了更多的选择。在线音乐要想赢得用户的喜爱，就必须从产品的体验入手，为消费者提供符合消费者不断提升的需求。网易云音乐一直秉承着让用户更好地发掘和共享优质音乐的理念，通过网络来促进中国音乐行业的发展。

根据网易云音乐的市场动向和大数据分析，未来的音乐产业将会有以下几个发展方向：第一，原创音乐由小众转向大众；二是互联网音乐进入了付费时代，越来越多的人选择了版权作品；三是分众化时代，音乐审美呈现出多元化、音乐形态百花齐放的倾向；四是移动互联网的发展，音乐平台将成为多种文化的传播场所，如音乐、影视、娱乐等。

5 大数据技术在音乐发展模式上的主要问题和未来展望

在大数据时代，音乐的发展面临的最大问题就是版权问题，免费和自由是互联网音乐所倡导的，用户可以通过互联网下载自己喜爱的歌曲，而不需要任何版权，就可以免费下载，这对整个音乐产业都产生了极大的冲击。大量的山寨产品，侵害了原作者的合法权益，导致了大量的盗版，对原作者的创作热情，也造成了极大的影响。传统唱片公司主导的音乐产业模式受到极大的挑战，长期来看，这对在线音乐产业的发展也是不利的^[16]。因此，必须从大数据技术层面入手，对音乐版权进行保护，保护音乐原创者的权益，保证歌曲的质量及产出数量，从而促进数字音乐的发展。利用大数据细化用户需求，精准施策^[17]。

网易云音乐一直秉承着让用户更好地发掘和共享优质音乐的理念，并通过网络来促进中国音乐行业的发展。现在，大数据技术已经被广泛的运用到了音乐搜索、音乐识别、音乐推荐、个性化定制等领域^[18]。迅速崛起，发展至今，推荐系统已经有 20 多年的历史^[19]。互联网时代的发展，音乐产业市场也当与时俱进^[20]。与传统音乐形式的音乐相比，网络音乐娱乐平台让使用者能够接触到更多的音乐人和艺人，探索新的音乐和音乐的衍生物，并与其它音乐爱好者产生共鸣和交流。随着手机网络的持续普及，在线娱乐需求不断增加，版权意识不断提高，中国在线音乐娱乐市场的用户数量不断增加。

6 结论

本文以网易云音乐为主要研究对象，结合大量的音乐及用户资源，使用大数据技术对网易云音乐进行数据分析，从数据收集、数据存储、数据分析、数据展示四个方面对数据进行处理。通过对用户活跃时段、歌手、歌曲、音乐评论话题的分析，深入了解用户行为以及音乐内容，通过音乐推荐，方便用户根据喜好选择音乐类型。

论文基于大数据分析，设计一个小型的音乐推荐系统，通过网络爬虫对数据进行收集，通过 Spark 对数据进行分析，通过 MongoDB 和 MySQL 对数据进行存储，最后通过 Spring Boot 和 Vue 设计一个后台管理系统，将所分析的数据进行可视化展示。

在本次设计中，主要通过对用户行为、音乐评论、歌手及歌曲进行一个多维度的数据分析，实现向用户推荐的功能。根据数据分析，我国在线音乐用户呈现出明显的年轻化特性，且覆盖用户广泛，全国各地用户均有触达。用户活跃时段主要集中于夜晚时段，熬夜人群较多。在线音乐消费趋向多元，音乐风格多样。

7 未来展望

音乐个性化推荐系统提升了用户的体验感，加强了用户之间的互动，可以根据用户的行为习惯个性化推荐，旨在帮助用户找寻高品质音乐。利用系统资源及优势使用大数据技术对用户进行个性化的推荐，将更大程度地促进音乐产业的发展。

本次设计完成了基本的数据分析，但仍然存在部分问题，目前爬取数据的平台没有提供开放数据，数据产生的爬文资料是通过线上爬虫的方式，但由于对方平台的技术较为繁杂多变，暂时没有办法及时快速的在短时间内更新获取最新的数据。

未来如果能够通过爬虫的方式快速取得网站数据，或者爬取数据的平台能够提供开放数据，将达到对最新资讯及时进行分析。如果能将这个计划推广到更多不同领域，比如影视领域、直播领域、抖音等多个领域，推荐系统将会有更多的发展。

参考文献

- [1]马越. 基于可视化技术的音乐数据分析平台的研究[D]. 北京邮电大学, 2021.
- [2]郭博林. 基于大数据分析的音乐个性化推荐系统应用研究[D]. 电子科技大学, 2018.
- [3]谭学清, 何珊. 音乐个性化推荐系统研究综述[J]. 现代图书情报技术, 2014(09):22-32
- [4]李丹丹. 音乐社交: 网易云音乐的互动仪式链研究[D]. 西南大学, 2016.
- [5]林菲. 网易云音乐社交化研究[D]. 山东师范大学, 2020.
- [6]方媛媛. 音乐类专业平台短视频运营方式研究[D]. 山东大学, 2021.
- [7]赵冬梅. 互联网音乐 APP 的创新之路[J]. 传媒, 2017(01):61-62.
- [8]吴吉义, 李文娟, 黄剑平. 移动互联网研究综述[J]. 中国科学:信息科学, 2015, 45(01):45-69.
- [9]任明飞, 李学军, 崔蒙蒙. 基于 MongoDB 的非关系型数据库的设计与开发[J]. 电脑知识与技术, 2019, 15(34):1-2.
- [10]胡于响. 基于 Spark 的推荐系统的设计与实现[D]. 浙江大学, 2015.
- [11]杨慧. 基于校园资源云的 Spark 图书推荐技术的研究[D]. 西安科技大学, 2017.
- [12]张天翔. 基于 NLP 和爬虫的在线文章阅读系统设计与实现[D]. 南昌大学, 2020.
- [13]迟美琪. 网易云音乐品牌传播现象研究[D]. 吉林大学, 2020.
- [14]黄珍珠. 大数据背景下数字自我的个性化呈现——基于在校大学生对网易云年度听歌报告体验的调查[J]. 科技传播, 2021, 13(12):130-133.
- [15]许炜, 周莉. 从网易云音乐排行榜看流行音乐的发展现状[J]. 艺术评鉴, 2021(04):166-168.
- [16]司凤朝, 杨逢柱. 大数据时代中国音乐产业机遇与挑战研究[J]. 艺术教育, 2020(06):63-66.
- [17]秦雅萌. 音乐虚拟社区用户黏性影响因素研究[D]. 河北大学, 2021.
- [18]李晨. 大数据时代背景下的数字音乐精准营销[J]. 传媒论坛, 2022, 5(02):56-58.
- [19]推荐系统实践[M]. 人民邮电出版社, 项亮, 2012
- [20]王小梅. 互联网时代音乐产业市场营销策略研究[J]. 中国产经, 2021(20):40-41.

附录 1

表 appendix-1 cloud_music_artist

```
CREATE TABLE `cloud_music_artist` (  
  `artist` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `id` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `title` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL  
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;  
--
```

表 appendix-2 cloud_music_comments

```
CREATE TABLE `cloud_music_comments` (  
  `user_name` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NULL,  
  `user_id` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NULL,  
  `user_age` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NULL,  
  `user_gender` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NULL,  
  `user_city` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NULL,  
  `user_introduce` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NULL,  
  `comment` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NULL,  
  `comment_id` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NULL,  
  `praise` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NULL,  
  `datetime` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NULL,  
  `song_id` text CHARACTER SET utf8mb4 COLLATE utf8mb4_bin NOT NULL  
) ENGINE = InnoDB CHARACTER SET = utf8mb4 COLLATE = utf8mb4_bin ROW_FORMAT = Dynamic;
```

表 appendix-3 cloud_music_soaring

```
CREATE TABLE `cloud_music_soaring` (  
  `song_id` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `name` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `url` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL  
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

致谢

本次毕业设计的研究和论文的撰写工作是在曹永忠老师的悉心指导下完成的，我非常感谢我的指导老师曹永忠老师，从起初题目的选择，资料的收集，系统的设计，功能的完善和最终论文的撰写都经过了导师的严格把关。本论文从初期选题、中期考核到最终成型，我的指导老师给予了我大量的指导意见，老师在忙碌的工作中，抽出时间来审查、修改我的论文，每一步都在导师的指导下完成，倾注了导师大量的心血。在此我向我的导师曹永忠老师表示衷心的感谢！

回首大学四年，在这四年里，我过得非常开心，这对于我而已是非常宝贵的四年时间，在这四年里，我结交了很多的好朋友，一起留下了很多令人难以忘怀的美好回忆。我非常感谢我的朋友们，一直包容我、鼓励我、帮助我，他们让我的大学校园生活丰富而精彩，让我成为了一名快乐的女大学生。

我也非常感谢作为学生干部的这段宝贵经验，作为一名学生干部，我学到了很多课堂以外的东西，懂得了责任与担当。在学生会的三年，我们举办了很多场活动，每一场活动对我而言都意义非凡，从干事到主任再到主席，这段历程让我成长了许多，这一切都让我成为了一个更好的自己。

而成为一名中共预备党员，这是令我非常自豪的一件事。入党的过程艰辛漫长，在这段时间里，跟随党的脚步，我不断检验自己、提升自己、完善自己。感谢党组织对我的培养与教育，让我可以成为一个更优秀的自己。

作为父母的孩子，我真的非常感谢我的父母，感谢他们把我抚养成人，给我提供好的成长环境。每次和他们打电话，爸爸妈妈总是会关心我吃的好吗，过得开心吗，什么时候可以回家。我的父母教育我，要多为他人着想，不可以成为一个自私的人。要学会和他人分享，要学会宽容，要做一个善良的人，要多帮助有困难的人。但也一定要做一个开心的人，爸爸总是对我说，要阳光一点，要开心一点。谢谢我的父母，让我可以成为一个很有底气的人，因为无论我在哪，我都知道在我的身后有最爱我的父母在一直支持着我。

四年前，我满怀期待来到闽江学院。离开时，我也依旧保持初心，立下新的愿景，不求最大但求最优，但求适应社会需要，成为一个有用的人。同时，希望大家身体健康，万事如意，前程似锦，一起成为更好的自己！