闽江学院

本科毕业论文（设计）

题　　目	基于大数据的在线教育问题数据分析
学生姓名	谢雯瑶
学　　号	3187103112
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
指导教师	曹永忠
职　　称	助理教授
完成日期	2022 年 4 月 29 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）《 基于大数据的在线教育问题数据分析 》，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）：谢雯瑶

2020 年 4 月 29 日

摘　　要

随着网络技术的不断发展，人们接受教育的方式也越来越多样化，近几年许多线下课程转成线上教学，网络教学越来越蓬勃的发展。如今，线上教学还存在着教学品质好坏的问题，线上教学对于面对面教学的最大问题在于品质不一；所以能够收集更多的资料来客观的表达线上教育的问题和特性将是一个非常重要的课题。

通过在开源数据网站下载在线教育行业的数据，对数据进行清洗、存储分析，将用户地区分布、用户类型分布以及课程分析结果进行了可视化展示，更加直观的了解在线教育的现状，根据用户自身的需求，从课程、城市等方面做出相应的学习规划时提供一份可供参考的数据。

通过本系统，能够把数据正规化，建立之间关联性，进行可视化展示；建立一个在线教育问题数据分析平台，相信能够让用户获得更多线上教育客观的资料。

关键词：在线教育；大数据平台；数据分析；数据可视化

Abstract

With the non-stop and rapid development of Internet technology, the way of education for people is also more and more diversified in recent years. The online education is also gradually rising under current situation of the development of online education enterprises lagging behind with the analysis of online teaching data, which can effectively solve the existing problems for online education. Online education problems are crucial for applying big data technology to upgrade its contain and availability.

With following processing including downloading the data of the online education on the open source data website, cleaning and data storages, visualization displaying the user regional distribution, user type distribution and course analysis results, more intuitively understanding the current situation of online education, and providing a copy of the data for reference when making the corresponding learning plans from courses and cities according to the requirements of users.

From the data results visualization, it can be visualized that the distribution of online education users geographically, the amount of users is gradually increasing, the demand for users is increasing, and it is great significance to recommend courses to users.

Key words: Online Education; Big Data Platform; Data Analysis; Data Visualization

目　　录

基于大数据的在线教育问题数据分析

谢雯瑶

（闽江学院 数学与数据科学学院（软件学院），福建 福州 350108）

1 绪论

在线教育是基于互联网的新教学方法，即一种利用计算机信息技术与网络手段，获取与共享知识信息、高效学习的全新教学方法；借助互联网，老师们能够不受空间约束随意地管理课堂,这对于以往的传统教育模式来说，是一个颠覆性的教育变革，正由于这种的变化，不断的变革着我们的教学环境和人生。在线课堂是指一种完完整整的课程，内容包含线上教学资源发布、资料的收集和使用、网络教室、老师答疑、与同学交流。

1.1 研究背景及意义

1.1.1 研究背景

在线教育是教育发展的新的方向和动力源泉，将会为我们带来教育的理念、制度、教学模式以及人才培养模式的深刻变革[1]。在线教育是利用互联网技术与信息技术，将线下课堂转移到线上课堂。在“互联网+”的今天，教育和学习越来越具有泛在、自主、自在的特点，在线教育是“互联网+”的泛网络教学，它强调了利用现代教育技术，尤其是因特网技术的作用，提倡人人都能学、随时随地学[2]。在2019年底发生的新冠肺炎疫情把我国的在线教育事业推向了一个新的高峰，全国各地的学生都开始向线上学习，所以各种创新的技术和运作方式得到了实践检验[3]。

随着因特网技术的迅速发展，近年来，网上教学逐步进入公众视线，4G和宽带的普及,中国的线上教学特别是手机网络教学正在蓬勃发展[4],现在以上网课与线下课相结合的方式，因此"互联网+教育"已经是当今教育的模式[5]。在线教育行业当前处于“一半是海水，一半是火焰”的状态。根据中国互联网信息中心（CNNIC）第46次公布的《中国互联网络发展状况统计报告》，该报告指出，截止到2020年，中国互联网用户数量达到了3.81亿，占到了总网民的40.5%。目前，国内出现了许多网络教育公司缺乏有效的盈利方式，很难在网络环境下发展；同时，随着传统的线下教学模式的逐步成熟，学生的需求增多，网络教学面临着学生的流失，在线教育企业在教育市场的占比可能会有所下降，造成公司发展停滞不前，乃至倒闭[6]。因此，如何利用大数据促进在线教育的

发展，为用户提供客观可供参考的数据，就成了当前在线教育事业发展的一个重大课题。

1.1.2 研究意义

网络与教育的深度结合导致了在线教育的诞生，从而引发了个性化教学，并由此形成了“人机结合”的教学模式[7]。网络教学、慕课等各种教学模式层出不穷，各种网络教学平台、网络教学应用等也随之出现。特别是在今年春季，由于国内新冠肺炎疫情的影响，教育部“停课不停学”的号召，用户对于线上课程的需求越来越多，所以网上教学平台已成为“互联网+教育”的一个重要阵地。所以，为了更好地服务于网络用户，本文对学习进度、用户流失率、推荐学习、回流率、登录地点分布等方面进行了深入的分析，从而为教育事业的发展提供数据参考服务有着重要的意义。

1.2 国内外研究现状

1.2.1 国内学者关于在线教育研究现状

管佳，李奇涛（2014）基于对中国网络教学的发展状况的研究，得出我们应该积极地掌握网络教育的发展趋势，并结合学者的兴趣和喜好，为我国教育信息化建设提供参考，这对教育信息化来说是既是挑战更是机遇[8]。吕海燕等（2017）基于计算机基础信息导学平台的数据分析，得出要引起学员对课程的重视，课程就应学员的兴趣爱好以及课程资源的自身特色[9]。张曦（2020）在冠状病毒肺炎疫情背景下，得出在线教育信息化的发展，同时也促进了中国教育事业快速发展[10]。郑勤华等（2020）基于疫情防控背景下对疫情时期的学生进行分析，由于在开学后的复课时间里，学校的课程安排相对较少，学生可以根据自己的兴趣爱好自主选择在线教学平台进行学习[11]。汤艳慧（2021）对在线教学平台进行分析，得出，线上教育得以发展，应充分利用大数据的优势，对教学平台数据进行分析和处理，来发挥在线教育的应用价值[12]。

1.2.2 国外学者关于在线教育研究现状

Junjing Zhao（2020）基于共享经济背景下，得出在线教育充分利用互联网信息技术，创造了多种多样的教学模式充分，促进学生的学业进步，促进了学生的学习和发展[13]。Anna Liu, Rui Zhang（2021）以猿辅导为研究对象，得出如果平台要实现长期稳定的发展，需要很好的满足需求，对于教育需求，只有通过提高产品质量和用户满意度体验，可以获得用户支持并产生用户价值；向潜在客户推荐产品，减少客户流失[14]。Oksana

Pirogova 等（2021）在俄罗斯的网络教育基础下，得出了在线教育市场已逐步扩大，且在线教育平台可以为学生提供全面便捷学习所需的功能，让用户在线上有个良好的体验[15]。Chen Lujun（2021）在大数据的背景下，针对不同的学生特点，制定相应的教学策略，让他们得以参加合适的培训，并能更好地理解他们的需求。通过实验研究，发现在大数据环境下，个性化网络教学平台有了显著的改进；从具体来看,在人工智能的背景下,线上用户数量增加了 9 %[16]。

1.3 研究内容与创新点

1.3.1 研究内容

本文主要是对在线教育资料进行数据分析，以在线教育资料为研究对象，对已有数据进行整理、分析，对在线教育事业的发展提供现实意义。主要目标是通过了解在线教育用户对线上课程学习的现状，从而在线教育企业可以根据现有的数据，改进它们的盈利方式，从而推动在线教育事业的发展。

首先分析在线教育的研究背景、意义、国内外研究现状以及研究内容与创新点。其次，介绍了此次数据分析所运用到的大数据平台及相关技术。再者，对此次系统进行需求分析以及系统的分析与设计，为之后系统实现奠定了坚实的基础。最后，对分析的结果进行罗列展示。论文架构如图 1-1 所示：

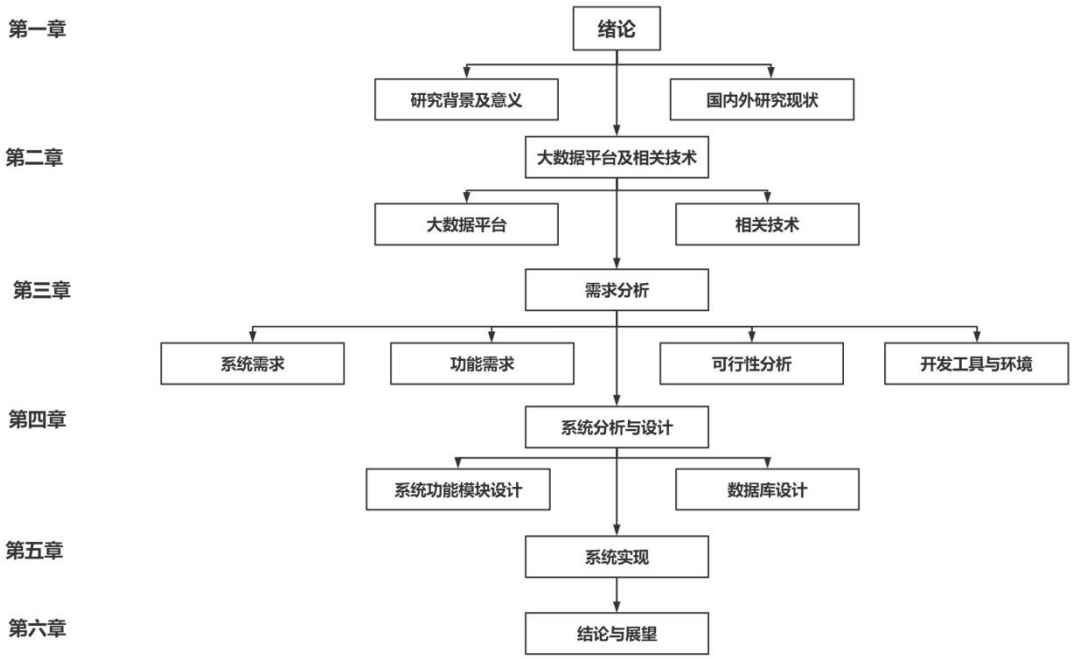

图1-1 论文架构图

1.3.2 创新点

本文研究的创新点在于运用数据分析来研究在线教育现状的创新问题，对在线教育的数据进行深度层次分析，如学习进度、用户流失率、推荐学习、回流率、登录地点分布这五个维度进行数据分析，通过可视化技术来展现在线教育的现状问题，对用户所学习的课程以及相似用户所学习的课程进行课程推荐，为用户提供客观、真实的参考资料，为在线教育盈利模式创新提供经验借鉴。所以将数据分析运用在在线教育盈利模式的创新上，大大丰富了该行业发展的模式。

2 大数据平台及相关技术

2.1 大数据平台

当前，大数据有三大优点：海量信息、处理速度快、数据种类繁多。我国大数据起步较晚，大数据的应用与发展尚处在起步阶段，但大数据在我国的重要性和应用价值却受到了极大的关注[17]。大数据平台是用来处理当今社会日益增长的数据，是一个用于存储，操作和显示的平台，它可以让开发者在“云”中运行。就像现在很多的舆论监控软件。大数据分析网络平台,是集合数据分析接入、处理、数据分析、储存、查询搜索、大数据分析挖掘等、应用接口等于一身的综合网络平台。在应用大数据平台时，它具有两个主要特征：数据多，数据之间有着很强的关联性[18]。

2.2 相关技术

2.2.1 数据库

（1）MongoDB 数据库

MongoDB 是由 MongoDB Inc 公司所开发的，它是一个介于关系数据库和非关系数据库之间的产品，它能够去存储更多更复杂的资料。MongoDB 最突出的特点是其支持的查询语言非常强大，其语法类似于面向对象的查询，它可以处理大部分的单一表单查询，也可以为数据建立索引，我们可以在 https://www.mongodb.com/try/download/community 下载此软件。

（2）MySQL 数据库

MySQL 由瑞典 MySQL AB 公司开发，属于 Oracle 旗下产品；是一种基于关系的数据库，它可以在不同的表格中存储数据，从而大大提高了数据的处理速度和灵活性。MySQL 具有许多优点，如速度快、成本低等优点；因此许多中小网站将其作为其站点的数据库；我们可以在 https://www.mysql.com/cn/downloads 下载此软件。

2.2.2 Spark 介绍

Spark，是一种通用的大数据计算框架，是加州大学伯克利分校 AMP 实验室（Algorithms, Machines, and People Lab）开发的通用内存并行计算框架，可以在大规模的、低延迟的数据分析应用程序使用。它存在着以下优点：一是易用性好，可以进行交互式编程；二是通用性好，提供完整强大的工具，如查询 SQL、机器学习等；三是随处

运行，可在多种情况下运行。

2.2.3 相关框架介绍

（1）Springboot Boot 框架

Spring Boot 是一个由 Pivotal 团队创建全新的架构，是一个基于 Java 的开源框架，用于创建微服务；它用于构建独立的生产就绪 Spring 应用。他为开发人员提高了生产力，缩短开发时间，同时 Spring 应用更容易进行理解和开发。

（2）Mybatis-plus 框架

Mybatis-plus 国人团队苞米豆在 Mybatis 的基础上开发的框架，这是 MyBatis 的一款增强工具，其目的是为了简化开发和提高效率，仅对 MyBatis 进行改进。它有依赖小、损耗小、支持代码生成、预防 SQL 注入，支持代码生成等优点。

（3）Vue 框架

Vue 框架是作者在 2013 年尤雨溪为了解决在大量的原型设计上，更快速的得到有形的东西，就创作了 Vue；这是一组 JavaScript 框架，用来建立用户接口；同时它具有许多优点，比如体积小，运行效率高，生态丰富学习成本低等。

2.2.4 ECharts 图库表介绍

ECharts 由百度团队开源的，并于 2018 年初捐赠给 Apache 基金会；它是基于 JavaScript 的数据可视化图表，提供了一个直观，互动，个性化的数据可视化图表，其中包括折线图、关系图、散点图等等。数据可视化可以让我们更快、更清晰、更直观的了解数据的变化趋势；我们可以在 https://echarts.apache.org/examples/zh/index.html 看到散点图、饼图、柱状图等图示例。

3 需求分析

3.1 系统需求分析

系统主要分为原始数据采集、数据预处理、数据分析、数据可视化 4 个部分。通过一系列操作，将数据可视化为直观便捷的柱状图、折线图等图表，为众多用户提供客观的资料以及在线教育平台的盈利方式具有重要意义。

3.2 功能需求分析

3.2.1 数据采集

数据采集，也叫数据获取，是对目标区域、场景的特定原始数据的收集，主要包括图像、文字、语音、视频等。首先在开源数据网站采集数据，要首先明确采集的目标数据，根据自身的需求进行数据的采集规划，获取相应的数据集。数据采集，就是将外部的数据和内部的数据进行连接，然后将所有的数据输入到系统中，如摄像机和话筒，就是一种收集数据的工具[19]。

3.2.2 数据预处理

在真实的世界里，数据往往是残缺不全的，肮脏的，无法直接进行数据挖掘，也不能让人满意。数据预处理技术是一种有效的方法，可以有效地改善数据挖掘的质量。从一个原始资料库到一个数据挖掘资料库，数据预处理技术是对数据进行加工[20]；利用这种数据处理技术，可以极大地改善数据挖掘模型的质量，缩短挖掘的时间。

3.2.3 数据分析

数据分析就是利用正确的统计分析方法，对海量的数据进行综合、理解、吸收，使其充分发挥其作用，是对资料进行细致的调查和归纳，以便从资料中抽取有用的资讯，并得出结论。

3.2.4 数据可视化

大数据可视化是指通过支持可视化的用户界面和支持分析过程的人机互动方法和技术，实现对数据挖掘的自动分析，将计算机的计算能力与人类的认知能力有效地结合在一起，从而直观了解大量的复杂数据[21]。

3.3 可行性分析

3.3.1 技术可行性分析

对于这次在线教育问题的数据分析系统的相关资料都可以在网上搜索、公开数据源网站获得，在开发过程中所需要的软件系统都可以通过相关网站免费获得，比如 python 软件可以在网站 https://www.python.org/自行下载所需的版本。

3.3.2 应用可行性分析

我们可以通过公开数据源来获取在线教育的数据。我们可以通过 Microsoft Edge 浏览器打开网址，注册社区账号，搜寻相关信息，就可以获取到相关的在线教育的数据，其中包括用户登录信息、学习信息、用户信息、用户地区切割信息以及用户距离等信息的情况，都是可靠真实的数据。

3.4 开发工具与运行环境

3.4.1 开发工具

如表 3-1、表 3-2 所示，本系统运用到的开发工具为 pycharm2021.3.3 及 IntelliJ IDEA 2020.3。

表 3-1 pycharm2021.3.3 相关资讯表

pycharm 2021.3.3 相关资讯
系统要求： ● Microsoft Windows 64、10 的 8 位版本 ● GNOME 或 KDE 桌面 ● macOS 10.13 或更高版本 ● 最低 2 GB RAM，建议 8 GB RAM ● 2.5 GB 硬盘空间，建议使用 SSD ● 最低 1024×768 屏幕分辨率 ● Python 2.7 或 Python 3.5 或更高版本

表 3-2 IntelliJ IDEA 2020.3 相关资讯表

IntelliJ IDEA 2020.3 相关资讯
系统要求： ● Microsoft Windows 10/8/7 (incl.64-bit) ● 4 GB RAM minimum, 8 GB RAM recommended ● 2 GB hard disk space + at least 1 G for caches ● 1024×768 minimum screen resolution ● JDK 1.6 or higher

3.4.2 运行环境

本系统运用到的运行环境为 Win10，其设备规格如下：

- 设备名称：LAPTOP-QQL2H806
- 处理器：AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz
- 机带 RAM：16.0 GB (15.4 GB 可用)
- 设备 ID：96693A90-9FF9-49D7-BDF8-F0B21E0A6C00
- 产品 ID：00342-36307-19027-AAOEM
- 系统类型：64 位操作系统, 基于 x64 的处理器
- Windows 规格如下：
- 版本：Windows 10 家庭中文版
- 版本号：20H2
- 操作系统版本：19042.630
- 图形卡：带有 WDDM 驱动程序的 Microsfot DirectX 9 图形设备

4 基于大数据的在线教育问题系统分析与设计

4.1 系统功能模块设计

4.1.1 系统总体模块

根据功能分析可以得到，系统总体模块如图 4-1 所示，本系统的功能是通过在开源数据网站获取在线教育的原始数据，根据原始数据的情况进行预处理，在对处理后的格式化数据进行多维度分析，并将分析后的结果，通过可视化技术将分析结果以直观的形式展示出来。

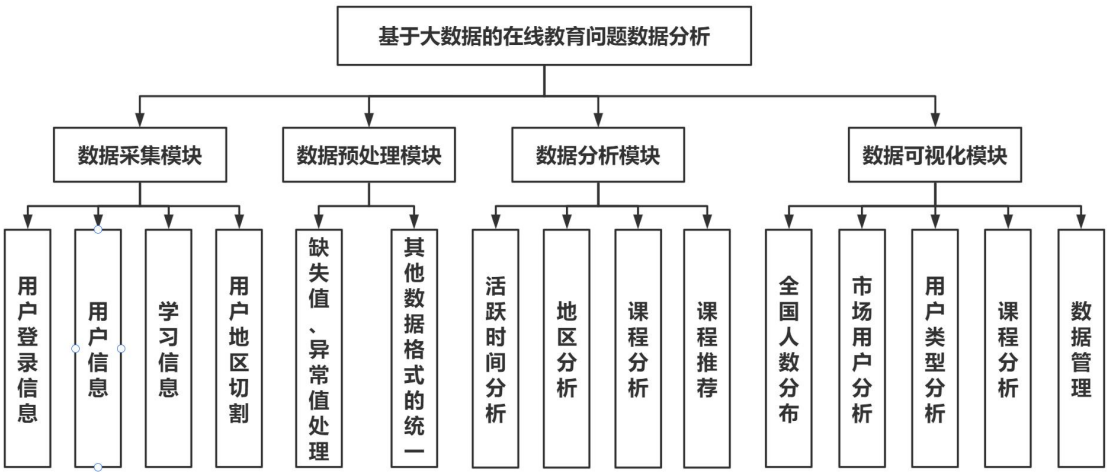

图 4-1 系统总体模块图

4.1.2 数据采集模块

数据采集模块图如图 4-2 所示，根据功能需求分析，开发者可以在现在的开源数据网站，下载现有的 CSV 数据文件，获取原始数据。

图 4-2 数据集获取流程图

4.1.3 数据预处理模块

如图 4-3 所示，开发者可以使用 Spark 对 CSV 文件进行预处理，对当中的缺失值或

空值进行填充或者过滤。把 CSV 表格处理为结构化数据，写入 MongoDB 中，部分 CSV 数据用于 Python 提取模型计算推荐课程。

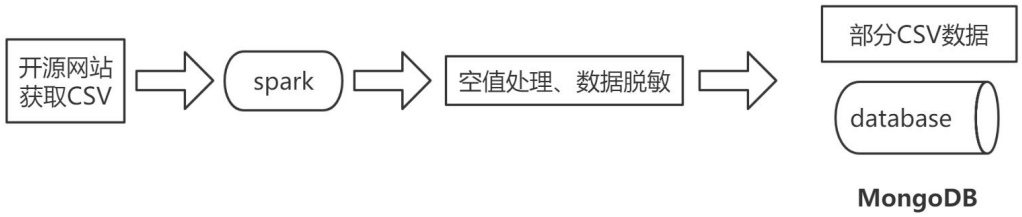

图 4-3 数据预处理模块图

4.1.4 数据分析模块

如图 4-4 所示，首先开发者根据范式建模规则，使用 Spark 对数据进行学习进度、用户流失率、推荐学习、回流率、登录地点分布这五个维度进行分析；部分 CSV 数据，提取模型计算推荐课程，将分析的结果数据存入 MySQL 数据库当中。

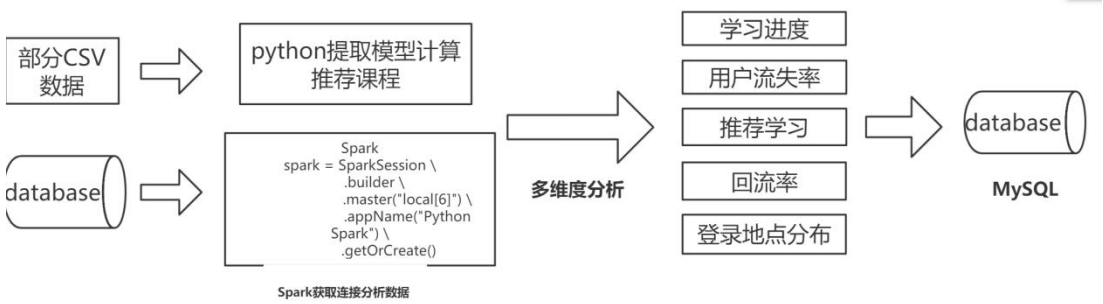

图 4-4 数据分析模块图

4.1.5 数据可视化模块

如图 4-5 所示，开发者编写后端 Java 代码使用现在的 Springboot 和 Mybatis-plus 框架对结果数据进行读取，并传送给前端 WEB。

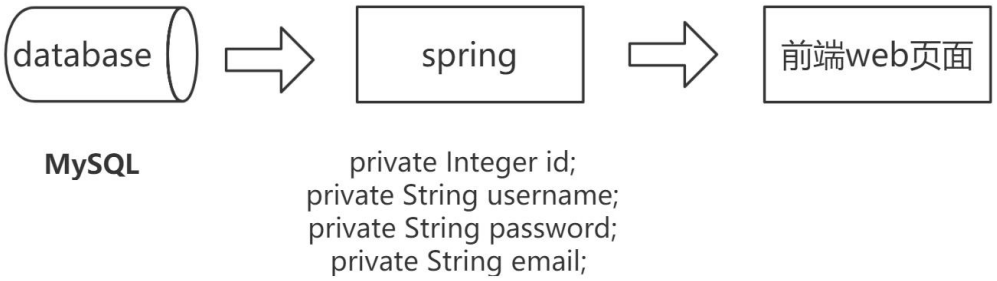

图 4-5 数据传递给前端流程图

如图 4-6 所示，开发者根据后端传递的 Json 信息，通过 Echars 将数据进行可视化分析。

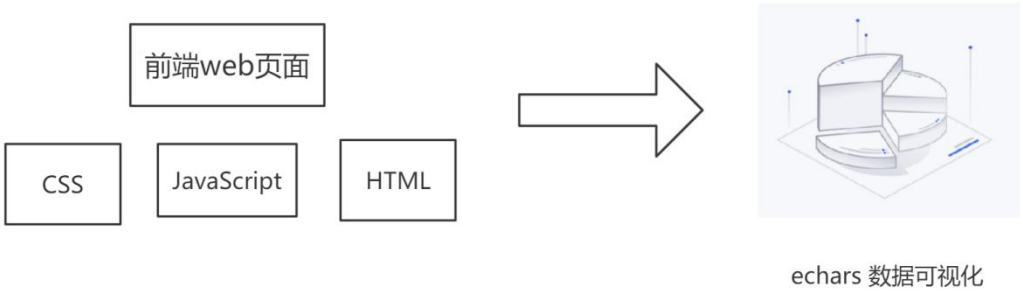

图 4-6 数据可视化流程图

4.2 数据库设计

4.2.1 数据库表关系

如图 4-7 所示，数据库资料表有 e_user、e_login、e_couse_popularity、e_loss_rate、e_course_complete、e_study_information、e_activity_time。

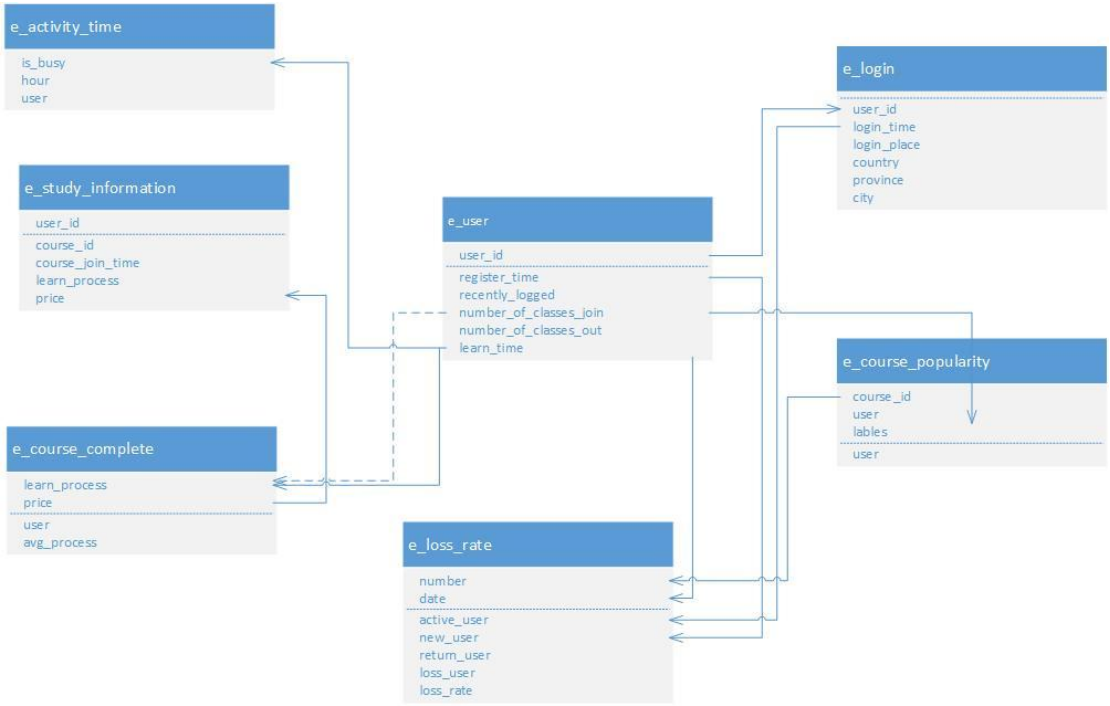

图 4-7 数据库表关系图

4.2.2 数据库资料表

根据系统功能需求分析，与系统的功能模块相结合的设计，对基于大数据的在线教育数据分析，数据需要 9 张数据表，每张数据表的名称及相关信息解释如下所示：

（1）e_login 资料表

表 4-1 e_login 资料表栏位规格书

栏位名称	型号	栏位解释
user_id	text	用户 id
login_time	text	登入时间
login_place	text	登入位置
country	text	国家
province	text	省份
city	text	县市

（2）e_activity_time 资料表

表 4-2 e_activity_time 资料表栏位规格书

栏位名称	型号	栏位解释
is_busy	text	是否是工作日
hour	int(11)	时间
user	bigint(20)	用户

（3）e_course_complete 资料表

表 4-3 e_course_complete 资料表栏位规格书

栏位名称	型号	栏位解释
price	text	价格
user	bigint(20)	用户
avg_process	double	平均进度

（4）e_course_popularity 资料表

表 4-4 e_course_popularity 资料表栏位规格书

栏位名称	型号	栏位解释
course_id	text	课程 id
user	bigint(20)	用户数

（5）e_loss_rate 资料表

表 4-5 e_loss_rate 资料表栏位规格书

栏位名称	型号	栏位解释
number	text	数量
date	text	日期
active_user	text	活动用户
new_user	text	新用户
return_user	text	老用户
loss_user	text	流失用户
loss_rate	text	流失率

（6）e_study_information 资料表

表 4-6 e_study_information 资料表栏位规格书

栏位名称	型号	栏位解释
user_id	text	用 id
course_id	text	课程 id
course_join_time	text	课程加入时间
learn_process	text	学习进度
price	text	价格

（7）e_user_distribution 资料表

表 4-7 e_user_distribution 资料表栏位规格书

栏位名称	型号	栏位解释
province	text	省份
user	bigint(20)	用户
count	bigint(20)	数量
avg_login	double	人均活跃次数
city	varchar(255)	城市

（8）e_user 资料表

表 4-8 e_user 资料表栏位规格书

栏位名称	型号	栏位解释
user_id	text	用户 id
register_time	text	注册时间
recently_logged	text	登记在案
number_of_classes_join	text	上课人数
number_of_classes_out	text	下课人数
learn_time	text	学习时间
PRIMARY id: OT NULL AUTO_INCREMENT COMMENT 'ID',		

（9）User 资料表

表 4-9 user 资料表栏位规格书

栏位名称	型号	栏位解释
id	text	用户 id
username	varchar(255)	用户名
password	varchar(255)	密码
nick_name	varchar(255)	昵称
age	int(11)	年龄
sex	varchar(255)	性别
address	varchar(255)	地址
avatar	varchar(255)	头像
account	decimal(10, 2)	账户余额
PRIMARY id: OT NULL AUTO_INCREMENT COMMENT ’ID’,		

5 基于大数据的在线教育问题系统实现

5.1 数据采集

我们可以通过开源数据网站来获取在线教育的 CSV 数据，通过 Microsoft Edge 浏览器打开网址，注册社区账号，搜寻在线教育信息，就可以获取到相关的在线教育的数据，其中包括用户登录信息、学习信息、用户信息、用户地区切割信息等数据集情况。

如图 5-1 所示，采集的 user 表数据包括 user_id、register_time、recently_logged、number_of_classes_join、number_of_classes_out、learn_time、school 相关信息。

user_id	register_time	recently_logged	number_of_classes_join	number_of_classes_out	learn_time	school
用户44251	2020/6/18 9:49	2020/6/18 9:49	0	0	41.25	
用户44250	2020/6/18 9:47	2020/6/18 9:48	0	0	0	
用户44249	2020/6/18 9:43	2020/6/18 9:43	0	0	16.22	
用户44248	2020/6/18 9:09	2020/6/18 9:09	0	0	0	
用户44247	2020/6/18 7:41	2020/6/18 8:15	0	0	1.8	
用户44246	2020/6/17 22:36	2020/6/17 22:36	0	0	48.92	
用户44245	2020/6/17 22:16	2020/6/17 22:16	0	0	0.18	
用户44244	2020/6/17 20:59	2020/6/17 21:34	0	0	0	
用户44243	2020/6/17 20:33	2020/6/17 20:34	0	0	297.47	
用户44242	2020/6/17 18:13	2020/6/17 18:13	0	0	0	
用户44241	2020/6/17 17:49	2020/6/17 17:49	0	0	0	
用户44240	2020/6/17 17:25	--	1	0	1667.28	
用户44239	2020/6/17 17:24	--	1	0	2109.75	
用户44238	2020/6/17 16:48	2020/6/17 16:48	0	0	263.2	
用户44235	2020/6/17 16:39	--	1	0	0	
用户44237	2020/6/17 16:39	--	1	0	10348.62	
用户44232	2020/6/17 16:39	--	1	0	9054.72	
用户44233	2020/6/17 16:39	--	1	0	11073.58	
用户44234	2020/6/17 16:39	--	1	0	4479.35	
用户44231	2020/6/17 16:16	2020/6/17 16:16	0	0	0.45	

图 5-1 user CSV 图

如图 5-2 所示可知，采集的 study_information 表数据包括 user_id、course_id、course_join_time、learn_process、price 相关信息。

user_id	course_id	course_join_time	learn_process	price
用户3	课程106	2020/4/21 10:11	width: 0%;	0
用户3	课程136	2020/3/5 11:44	width: 1%;	0
用户3	课程205	2018/9/10 18:17	width: 63%;	0
用户4	课程26	2020/3/31 10:52	width: 0%;	319
用户4	课程34	2020/3/31 10:52	width: 0%;	299
用户4	课程22	2020/3/31 10:52	width: 0%;	199
用户4	课程17	2020/3/31 10:52	width: 0%;	299
用户4	课程31	2020/3/31 10:52	width: 0%;	109
用户4	课程4	2020/3/13 11:15	width: 2%;	369
用户4	课程51	2020/3/5 16:50	width: 0%;	
用户4	课程56	2020/3/3 10:49	width: 0%;	299
用户4	课程132	2020/3/3 10:49	width: 0%;	199
用户4	课程171	2020/3/3 10:49	width: 0%;	299
用户4	课程48	2020/3/3 10:49	width: 0%;	299
用户4	课程87	2019/10/9 10:00	width: 0%;	49

图 5-2 study_information CSV 图

如图 5-3 可知，采集的 login_data 数据包括 user_id、login_time、login_place、国家省份、县市相关信息。

	Unnamed: 0	user_id	login_time	login_place	国家	省份	县市
0	0	用户3	2018/9/6 9:32	中国广东广州	中国	广东	广州
1	1	用户3	2018/9/7 9:28	中国广东广州	中国	广东	广州
2	2	用户3	2018/9/7 9:57	中国广东广州	中国	广东	广州
3	3	用户3	2018/9/7 10:55	中国广东广州	中国	广东	广州
4	4	用户3	2018/9/7 12:28	中国广东广州	中国	广东	广州
5	5	用户3	2018/9/10 9:18	中国广东广州	中国	广东	广州
6	6	用户3	2018/9/10 9:53	中国广东广州	中国	广东	广州
7	7	用户3	2018/9/10 11:28	中国广东广州	中国	广东	广州
8	8	用户3	2018/9/10 14:04	中国北京	中国	北京	
9	9	用户3	2018/9/10 14:36	中国广东广州	中国	广东	广州
10	10	用户3	2018/9/10 17:38	中国广东	中国	广东	
11	11	用户3	2018/9/10 18:17	中国广东广州	中国	广东	广州
12	12	用户3	2018/9/11 9:40	中国广东广州	中国	广东	广州
13	13	用户3	2018/9/11 14:38	中国广东广州	中国	广东	广州
14	14	用户3	2018/9/11 16:32	中国广东广州	中国	广东	广州
15	15	用户3	2018/9/11 17:00	中国广东广州	中国	广东	广州
16	16	用户3	2018/9/11 17:33	中国广东广州	中国	广东	广州
17	17	用户3	2018/9/12 10:39	中国广东广州	中国	广东	广州

图 5-3 login_data CSV 图

如图 5-4 所示，用户地区切割数据包括 user_id、login_time、login_place、时间差、国家、省份、地区相关信息。

	user_i	login_time	login_place	时间差	国家	省份	地区
0	用户3	2018-09-06	中国广东广州	651	中国	广东	广州
1	用户3	2018-09-07	中国广东广州	650	中国	广东	广州
2	用户3	2018-09-10	中国广东广州	647	中国	广东	广州
3	用户3	2018-09-10	中国北京	647	中国	北京	
4	用户3	2018-09-10	中国广东	647	中国	广东	
5	用户3	2018-09-11	中国广东广州	646	中国	广东	广州
6	用户3	2018-09-12	中国广东广州	645	中国	广东	广州
7	用户3	2018-09-13	中国广东广州	644	中国	广东	广州
8	用户3	2018-09-14	中国广东广州	643	中国	广东	广州
9	用户3	2018-09-18	中国广东广州	639	中国	广东	广州
10	用户3	2018-09-19	中国广东广州	638	中国	广东	广州
11	用户3	2018-09-20	中国广东广州	637	中国	广东	广州
12	用户3	2018-09-21	中国广东广州	636	中国	广东	广州
13	用户3	2018-09-23	中国广东广州	634	中国	广东	广州
14	用户3	2018-09-25	中国广东广州	632	中国	广东	广州
15	用户3	2018-09-26	中国广东广州	631	中国	广东	广州
16	用户3	2018-09-27	中国广东广州	630	中国	广东	广州
17	用户3	2018-09-28	中国广东广州	629	中国	广东	广州
18	用户3	2018-09-29	中国广东广州	628	中国	广东	广州
19	用户3	2018-09-30	中国广东广州	627	中国	广东	广州
20	用户3	2018-10-08	中国广东广州	619	中国	广东	广州

图 5-4 用户地区切割 CSV 图

5.2 数据预处理与存储

（1）数据预处理

开发者使用 spark 对原始数据文件进行预处理，对当中的缺失值或空值进行填充或者过滤。如图 5-5 所示，以 e_login 数据表格为例，用户登录信息的原始数据 use_id 为用户 3，经过 spark 预处理后，得到的 use_id 为 3；

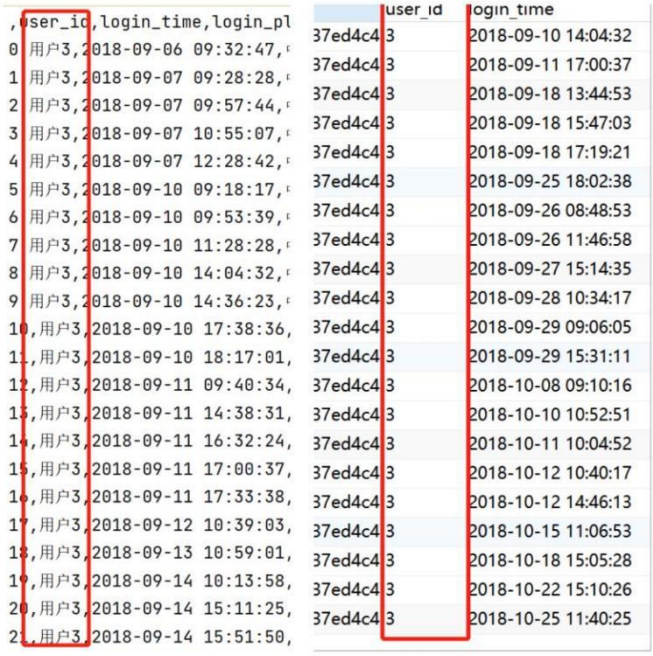

图 5-5 登录信息数据预处理图

e_login 数据预处理相关代码如下：

```
    login_date_new = spark.read.csv(r"./../DataSet/login_date.csv", encoding="Utf-8", header=True) \
    .select(
    F.regexp_replace(F.col("user_id"), "用户", "").alias("user_id"),
    "login_time",
    "login_place",
    F.col("国家").alias("country"),
    F.col("省份").alias("province"),
    F.col("县市").alias("city")
).distinct()
```

（2）数据存取

开发者在数据预处理后，得到格式化数据，将格式化数据写入到 MongoDB 数据库中；在进行数据分析前，读处 MongoDB 中的宽表数据，对格式化数据进行多维度分析之后存入到 MySQL 数据库当中。数据存取相关代码如下：

数据读取：

```
    def get_source(self):
    # 得到宽表数据
    return MongoUtils().read_from_mongo(self.spark, collection="e_login")
```

数据存储：

```
    MongoUtils().sink_to_mongodb(login_date_new, collection="e_login")
    MySQLUtils().sink_to_mysql(activity_analysis, table='e_activity_time')
```

5.3 数据分析与可视化

数据预处理完成后，我们需要对在线教育数据进行多维度分析，例如学习进度、用户流失率、推荐学习、回流率、登录地点分布等等。编写后端 Java 代码使用现在的 SpringBoot 和 Mybatis-plus 框架对结果数据进行读取传送 Json 给前端，使用前端 VUE 框架，Echars 图表库完成数据可视化。

5.3.1 全国人数分布

由图 5-6 可看出，上图中橙色小点代表着各个省份的省会城市，黄色小点的省会城市代表着登录用户人数较多的省份。同国家归属一样，我们将用户登录最多的地区作为其常用登陆地，且认为其为用户的归属地。我们能够图中看到，登入用户最多的省份城市是广东广州、贵州贵阳、湖北武汉、河南郑州、山东济南以及河北石家庄。

图 5-6 全国各省登入分布

5.3.2 市场用户分布

（1）用户登录平均分布

如图 5-7 所示，用户数以蓝色柱状图表示，登录次数以绿色柱状图表示。平均登录次数以折线图表示；从用户数和登录次数上看，广东省的用户数为 8981 人，登录次数为 12088 次，远远超过其他省份，其次为为湖北省；从平均登录次数上看，平均登录次数为登录次数与用户数之比，可知，广东省与贵州省的平均登录次数较高，均在 13 次以上；天津、香港、北京、上海、台湾的平均登录次数较低，均在 4 次以下。

	用户数	登录次数	平均登录次数
山西	828	887	10.7186
广东	8981	12088	13.4603
天津	462	180	3.9069
河北	1711	1470	8.5961
湖南	1509	1310	8.6832
辽宁	602	291	4.8455
福建	974	655	6.7331
浙江	1885	1336	7.0907
安徽	1084	1033	9.5314
香港	144	34	2.3681
云南	499	375	7.515
广西	1432	1405	9.8128
河南	1730	1855	10.7225
北京	2073	494	2.3859
贵州	1274	1878	14.7457
青海	66	45	6.8939
陕西	1278	1208	9.4585
四川	1697	1309	7.7189
甘肃	415	413	9.9711
江苏	1981	1123	5.6724
宁夏	165	148	8.9758
湖北	3049	3314	10.8721
黑龙江	472	377	7.9979
西藏	14	10	7.2857
上海	1511	536	3.5506
内蒙古	267	187	7.0037
山东	1843	1487	8.0705
重庆	1295	1316	10.1645
海南	230	206	8.9652
江西	902	579	6.4257
吉林	327	228	6.9878

图 5-7 用户登录平均分布及数据视图

（2）用户活跃时间分布

从图 5-8 上可以看出，从 0 点到 7 点，工作日和非工作日的活跃用户数都在 2570 以下，工作日活跃人数最低达到 192 人，非工作日活跃人数最低达到 90 人；8 点之后开始急剧增加，工作日活跃人数最高达到 10461 人，非工作日的最高活跃人数达到 5313 人，相比休息日的活跃人数之下，工作日的活跃人数偏多，这说明用户都比较喜欢在 8 点之后以及工作日的时间进行课程的学习。

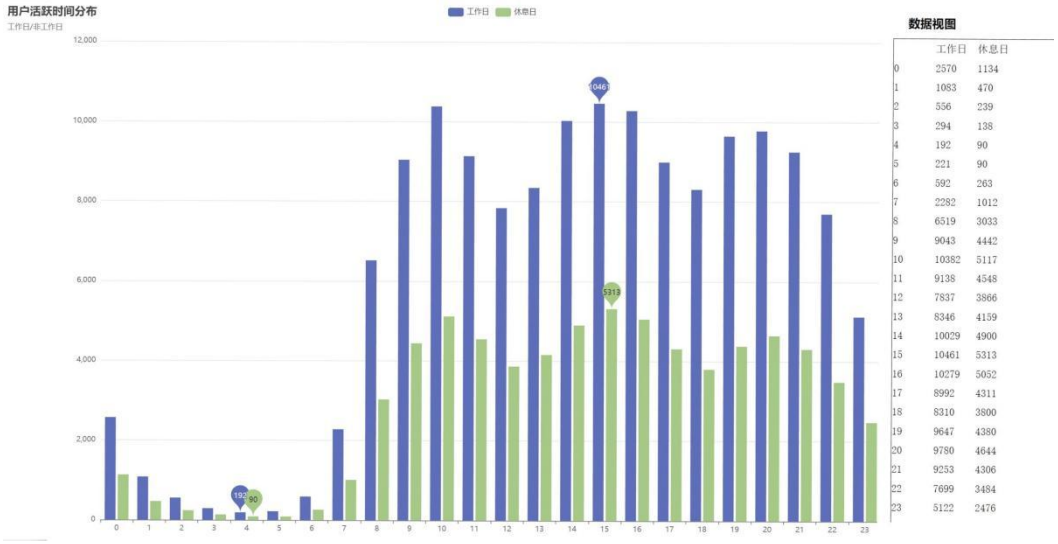

图 5-8 用户活跃时间分布及数据视图

5.3.3 用户类型分析

（1）活跃用户

如图 5-9 所示，活跃用户即当天有学习课程的用户；活跃用户变化趋势在 2020 年 2 月 6 日之前，相对于初期都是平缓的增加的，从 2020 年 2 月 6 日之后开始急剧增加，最高活跃用户数在 18024 人，这说明，由于 2020 年新冠肺炎疫情的影响，用户在线上学习的频率增加，活跃用户数也随着增加。

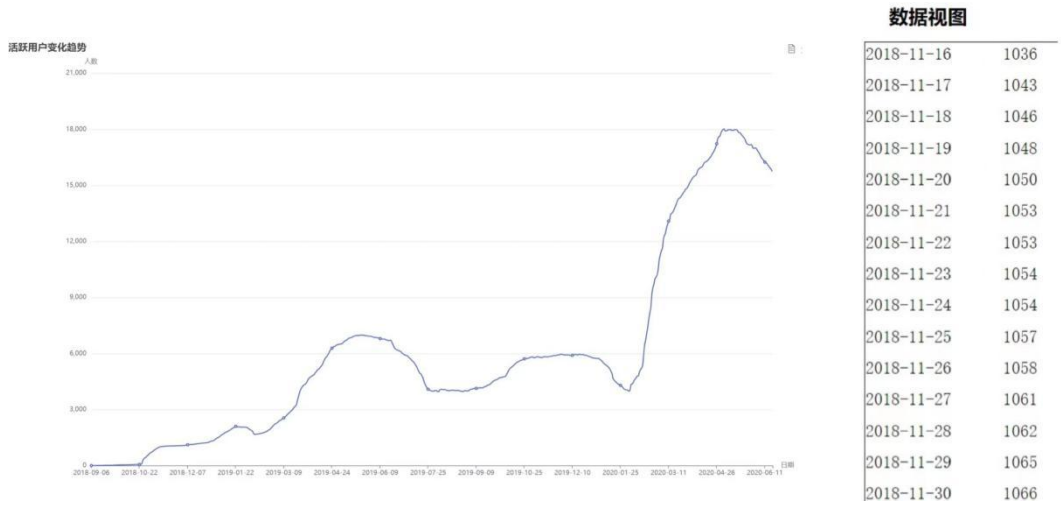

图 5-9 活跃用户变化趋势及数据视图

（2）新用户

如图 5-10 所示，新用户即当天注册的用户；在 2020 年 1 月 24 日之前，新注册用户人数约为 300 人以下，但在 2020 年 1 月 24 日到 2020 年 4 月 5 日之间，新注册人数急剧增加，新用户注册人数高达 390 人，并呈陡峭式增长。由此可知，在 2020 年 1 月 24 日到 2020 年 4 月 5 日这段时间内，用户对于线上课程的学习需求急剧增多。

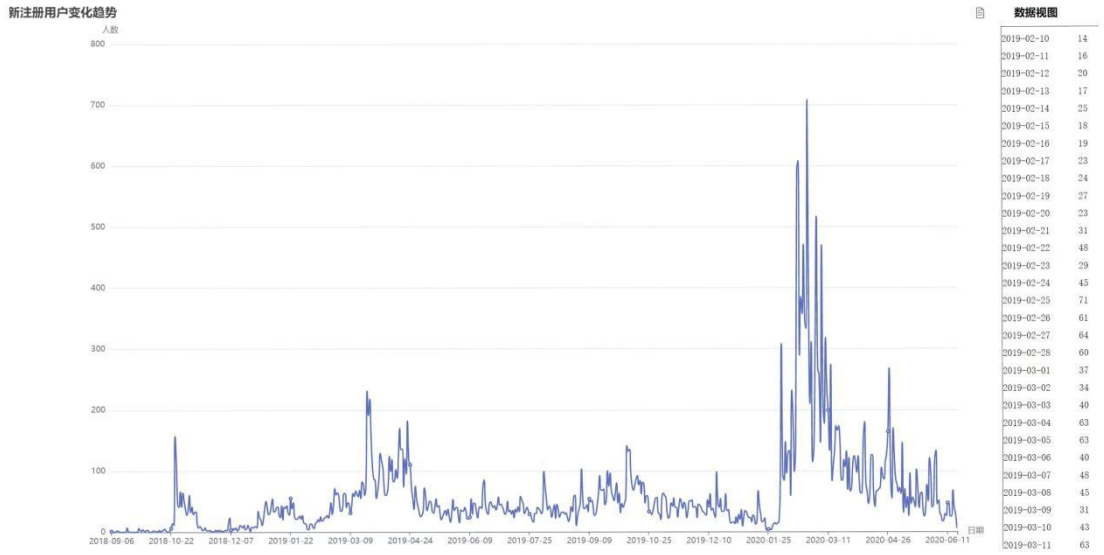

图 5-10 新用户变化趋势及数据视图

（3）回流用户

如图 5-11 可知，回流用户即隔 7 天后首次登录的用户；到 2020 年 1 月 24 日，回流用户人数约在 20 人之下，从 2020 年 1 月 24 日开始，回流用户急剧上升，回流用户数最高达到 53 人，这足以表示在 2020 年 1 月 24 日之后，人们对于在线教育的需求增加。

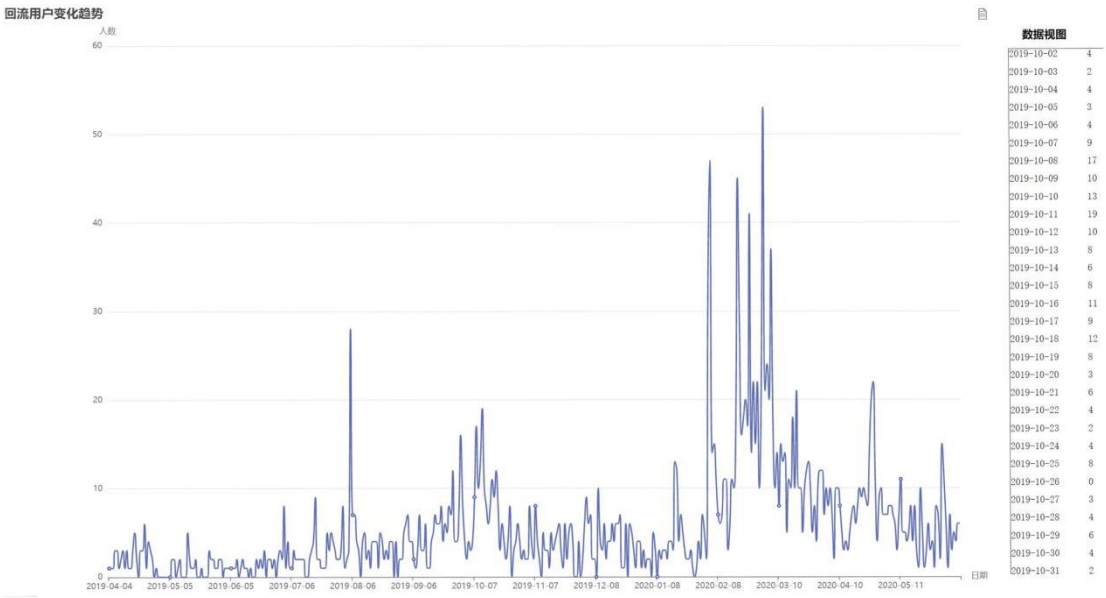

图 5-11 回流用户变化趋势及数据视图

（4）流失用户

如图 5-12 可知，流失用户即隔七天后没有登录的用户；在 2020 年 1 月 16 日到 2020 年 1 月 21 日这个区间，流失用户变化趋势较大，最高达到 313 人。从 2020 年 1 月 15 号开始流失率波动趋于稳定，最低达到 10 人。由此可以看出，经历新冠肺炎疫情爆发后，用户选择在线上接受教育学习的趋势状态趋于稳定。

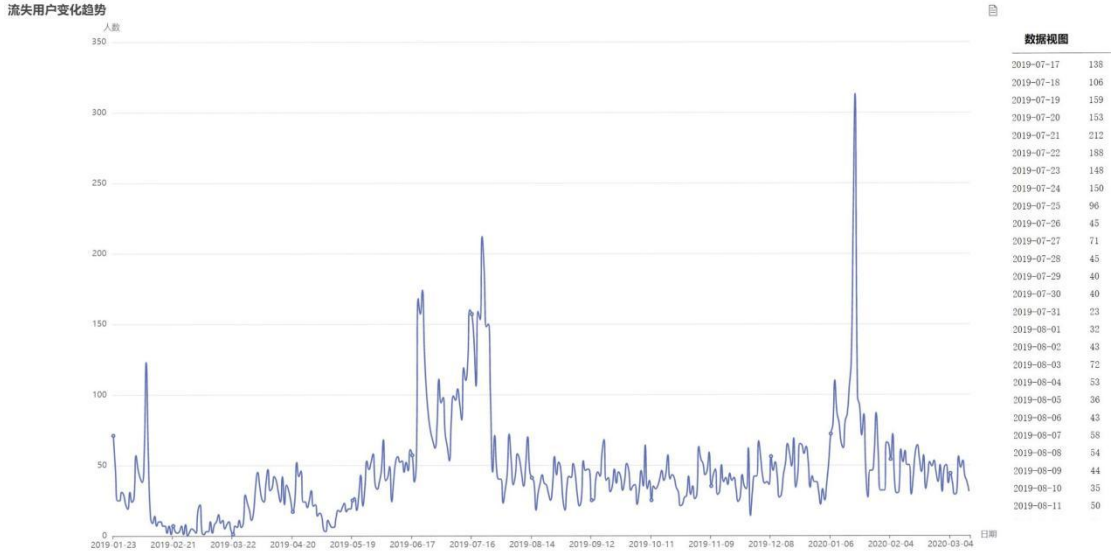

图 5-12 流失用户变化趋势及数据视图

（5）流失率

如图 5-13 所示可知，流失率即回流用户与流失用户之比为用户流失率；用户流失率在三个时间段的波动较大，分别是 2019.1.29-2019.3.6、2019.6.22-2019.7.28、2019.12.19-2020.2.29 这三个区间；相对于这三个时间段其他时间段而言，波动的幅度较小。

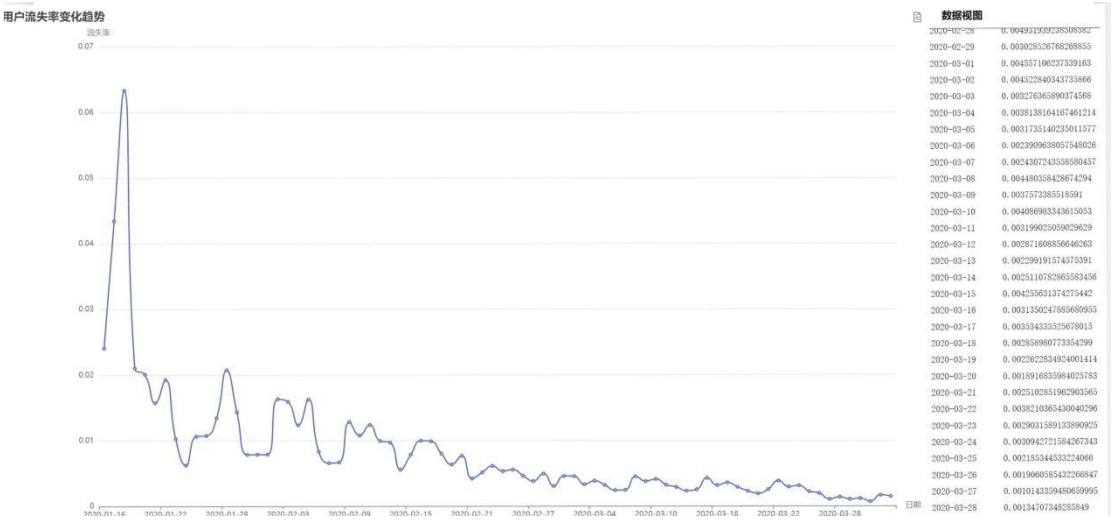

图 5-13 用户流失率变化趋势及数据视图

5.3.4 课程分析推荐

（1）各价位课程完成情况

如图 5-14 所示可知，蓝色柱状图为购买人数，绿色折线图为课程的完成比例，横轴为课程价格，从购买人数上看，价格在 0 元、109 元及 299 元的购买人数偏多，人数均在 10000 人以上，其中免费课程的购买人数达到 32342 人；价格在 229 元、800 元的购买人数在 100 人以下；从完成比例上看，价位在、169、700、800 上完成程度较高的，其中完成比例高达 85%；价位在 29 元、129 元、179 元、229 元、319 元、3000 元等 9 个价位的课程，完成比例均在 10%以下。

图 5-14 各价位课程完成情况及数据视图

（2）欢迎程度

如图 5-15 所示可知，课程 76、课程 31 是较为受欢迎的，人数都在 9000 人以上，其中课程 76 受欢迎人数达到 13265 人；相对课程 76、课程 31 的其他课程而言，课程的受欢迎人数都在 3000 人以上；这说明，线上课程对用户来说，比较便易和喜欢的。

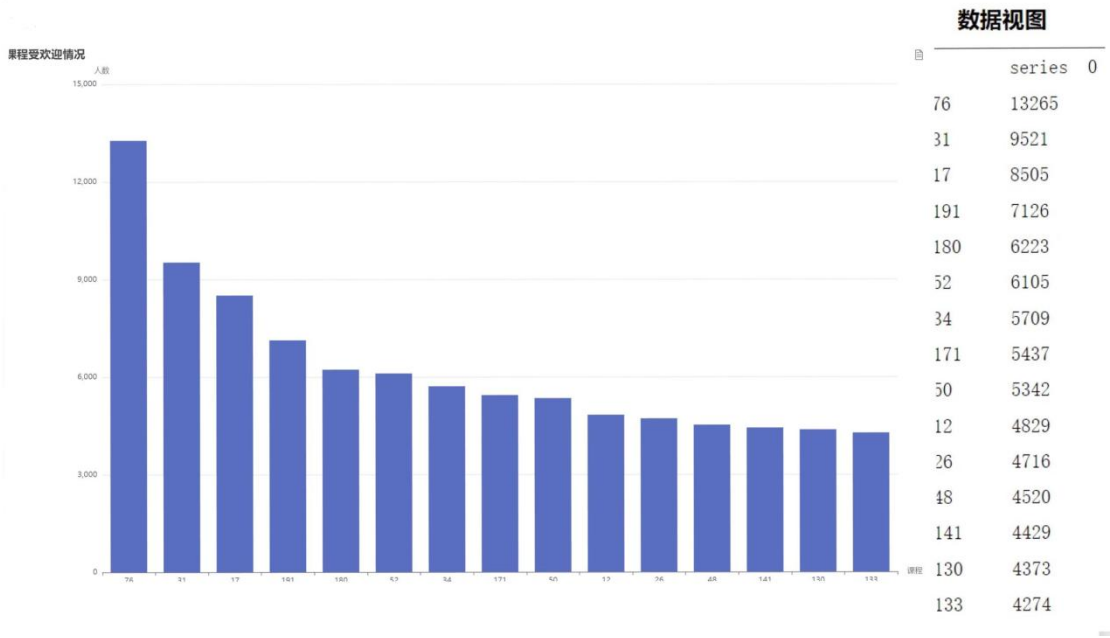

图 5-15 课程欢迎程度及数据视图

（3）课程推荐

如图 5-16 所示可知，可对总学习进度最高的 5 个用户，用户 ID 分别为 1193、13841、32684、36989、24985，对这 5 个用户的数据进行课程推荐；根据这 5 个用户所选的课程，对照其相似用户所选的课程进行线上课程推荐，从而已达到根据用户自己的需求，给予合适的建议。

线上课程智能推荐

用户ID	推荐课程1	推荐课程2	推荐课程3	推荐课程4	推荐课程5	推荐课程6	推荐课程7
1193	课程97	课程99	课程134				
13841	课程34	课程191	课程10	课程5	课程24	课程134	课程19
32684	课程24	课程53	课程115	课程19	课程240	课程76	课程38
36989	课程191	课程32	课程24	课程196	课程149	课程53	课程153
24985	课程191	课程32	课程5	课程196	课程24	课程175	课程153

图 5-16 课程推荐

5.3.5 数据管理

（1）用户管理

如图 5-17 所示可知，开发者可以对用户的 ID、用户名、昵称、年龄、性别、地址进行编辑、删除、新增、查询等操作功能。

图 5-17 用户管理

（2）个人信息

如图 5-18 所示可知，开发者可以对个人相关信息：用户名、昵称、年龄、性别地址进行修改及保存功能。

图 5-18 个人信息

6 结论以及展望

6.1 结论

本次设计主要是以开源数据网站的在线教育 CSV 文件为原始数据，对清洗后的数据进行客观的分析，建立起数据之间关联性，最后进行可视化展示；建立一个在线教育问题数据分析平台，能够让用户通过这个平台获取到准确的、真实的、实用的客观数据，从而推动对在线教育的发展。

本次系统通过对用户在线学习所产生的登录数据、学习进度、课程信息的数据进行采集、分析，对网络教学行为的分析，找出了网络教学行为的主要特点和影响因素。通过平台展示的客观数据上看，产生了以下信息：用户所学习的时间在工作日以及早晨 8 点后，用户的活跃人数相对较高的；同时对于价位相对较低的课程购买人数相对较多；再者新冠肺炎疫情爆发后，用户的活跃程度、新用户，回流用户的人数都在增加，流失用户的人数逐渐减少，流失率趋于平缓；因此想要让用户对课程引起足够的重视，就需要根据用户的兴趣爱好、以及自身的需求进行课程推荐，同时课程价格的合理度要满足大部分用户的需要。

用户可以根据此系统获取更加客观的线上教育的资料，让用户能够获取更正确、更合理、更物廉价美的、内容实在的线上教学课程。

6.2 展望

本文的研究尚有一些缺陷，收集的资料无法拿到最新的及时资料；线上教育平台未提供较多的资料在线上收集；由于时间上的关系，收集的资料还不够，样本数据的大小不一，数据的代表性和精确度，仍然存在一定的偏差。因此此次的设计还有很多缺陷有待完善，希望以后能够进一步了解到大数据的数据分析，进行更深层次的研究。

参考文献

[1] 刘盾. 在线教育的“烧钱争霸战”[N]. 中国教育报,2014-04-02(005).

[2] 朱新顺.“互联网+”时代在线教育研究与探索[J].现代信息科技,2019,3(22):146-147.

[3] 李准. 新东方在线科技控股公司 OMO 模式研究[D].吉林大学,2021.

[4] 蒋大成,王明宇.中国在线教育的现状和发展对策研究[J].电子商务,2015(09):68-69.

[5] 邵攸妮.互联网在现代教育的应用——在线教育的兴起与影响[J].中国多媒体与网络教学学报(上旬刊),2021(10):25-28.

[6] 邱宇,杜学元.试论在线教育企业面临的问题与对策[J].中国新通信,2021,23(22):161-163.

[7] 邢西深,李军.“互联网+”时代在线教育发展的新思路[J].中国电化教育,2021(05):57-62.

[8] 管佳,李奇涛.中国在线教育发展现状、趋势及经验借鉴[J].中国电化教育,2014(08):62-66.

[9] 吕海燕,周立军,张杰.大数据背景下教育数据挖掘在学生在线学习行为分析中的应用研究[J].计算技术与自动化,2017,36(01):136-140.

[10] 张曦.从在线教育临危上阵 谈谈教育信息化的“抗疫力”及发展建议[J].中国现代教育装备,2020(04):7-9.

[11] 郑勤华,秦婷,沈强,桂毅,周晓红,赵京波,王祎,曹一鸣.疫情期间在线教学实施现状、问题与对策建议[J].中国电化教育,2020(05):34-43.

[12] 汤艳慧.大数据在在线教学平台中的应用[J].办公自动化,2021,26(09):21-22.

[13] Junjing Zhao. Research on the Current Situation and Development Strategy of Internet Online Education Under the Background of Sharing Economy[P]. 2020 2nd Scientific Workshop on Advanced in Social Sciences, Arts & Humanities,2020.

[14] Anna Liu, Rui Zhang. Research on New Media Marketing Strategies of Ape Tutoring Online Education[J]. Scientific Journal of Economics and Management Research,2021,3(12)：131-136.

[15] Oksana Pirogova,Nataliia Temnova,Elena Markova. Online education in Russia: status and development trends[J]. E3S Web of Conferences,2021,258(3)：10020.

[16] Chen Lujun. Application of Artificial Intelligence Technology in Personalized Online Teaching under the Background of Big Data[J]. Journal of Physics: Conference Series,2021,1744(4)：042208.

[17] 黄陵.网络环境下的大数据采集和处理[J].网络安全技术与应用,2021(07):71-72.

[18] 刘一凡,李子乾,张月,宋灿,李冀.大数据平台的自动运维及监控技术[J].长江信息通信,2022,35(02):173-175.

[19] 杨玉.云计算架构下互联网大数据采集模型设计[J].电脑知识与技术,2019,15(05):19-20.

[20] 菅志刚,金旭.数据挖掘中数据预处理的研究与实现[J].计算机应用研究,2004(07):117-118+157.

[21] 任磊,杜一,马帅,张小龙,戴国忠.大数据可视分析综述[J].软件学报,2014,25(09):1909-1936.

附录 1 资料表 SQL 叙述

表附录 1 e_activity_time 资料表 SQL 敘述

```
DROP TABLE IF EXISTS `e_activity_time`;
CREATE TABLE `e_activity_time`  (
  `is_busy` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `hour` int(11) NULL DEFAULT NULL,
  `user` bigint(20) NOT NULL
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
--
```

表附录 2 e_course_complete 资料表 SQL 叙述

```
CREATE TABLE `e_course_complete`  (
  `price` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `user` bigint(20) NOT NULL,
  `avg_process` double NULL DEFAULT NULL
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
-
```

表附录 3 e_login 资料表 SQL 叙述

```
CREATE TABLE `e_login`  (
  `user_id` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `login_time` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `login_place` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `country` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `province` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `city` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

表附录 4 e_loss_rate 资料表 SQL 叙述

```
CREATE TABLE `e_loss_rate`  (
  `number` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `date` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `active_user` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `new_user` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `return_user` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `loss_user` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `loss_rate` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

表附录 5 e_study_information 资料表 SQL 叙述

```
CREATE TABLE `e_study_information`  (
  `user_id` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
```

```
  `course_id` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `course_join_time` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `learn_process` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `price` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

表附录 6 e_user 资料表 SQL 叙述

```
CREATE TABLE `e_user`  (
  `user_id` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `register_time` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `recently_logged` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `number_of_classes_join` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `number_of_classes_out` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `learn_time` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

表附录 7 e_user_distribution 资料表 SQL 叙述

```
CREATE TABLE `e_user_distribution`  (
  `province` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `user` bigint(20) NOT NULL,
  `count` bigint(20) NOT NULL,
  `avg_login` double NULL DEFAULT NULL,
  `city` varchar(255) CHARACTER SET utf8 COLLATE utf8_general_ci NULL DEFAULT NULL
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

表附录 8 e_course_popularity 资料表 SQL 叙述

```
DROP TABLE IF EXISTS `e_course_popularity`;
CREATE TABLE `e_course_popularity`  (
  `course_id` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,
  `user` bigint(20) NOT NULL
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

表附录 9 userSQL 资料表 SQL 叙述

```
CREATE TABLE `user`  (
  `id` int(11) NOT NULL AUTO_INCREMENT COMMENT 'ID',
  `username` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci NULL
DEFAULT NULL COMMENT '用户名',
  `password` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci NULL
DEFAULT NULL COMMENT '密码',
  `nick_name` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci NULL
DEFAULT NULL COMMENT '昵称',
  `age` int(11) NULL DEFAULT NULL COMMENT '年龄',
  `sex` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci NULL DEFAULT NULL
```

```
COMMENT '性别',
  `address` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci NULL DEFAULT NULL COMMENT '地址',
  `avatar` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci NULL DEFAULT NULL COMMENT '头像',
  `account` decimal(10, 2) NULL DEFAULT NULL COMMENT '账户余额',
  PRIMARY KEY (`id`) USING BTREE
) ENGINE = InnoDB AUTO_INCREMENT = 28 CHARACTER SET = utf8mb4 COLLATE = utf8mb4_unicode_ci COMMENT = '用户信息表' ROW_FORMAT = Dynamic;
```

附录 2 文献综述

1 前言

在线教育起源于美国，在 20 世纪 60 年代美国伊利诺伊大学将联网计算机引用在教学活动中。互联网教育不仅打破了时间与空间的壁垒，并开始通过更低的成本和多样化的教学模式与评估模式打动用户，满足用户个性化需求。中国互联网络信息中心(CNNIC)发布第 46 次《中国互联网络发展状况统计报告》报告显示，截至 2020 年 6 月，我国在线教育用户规模达 3.81 亿，占网民整体的 40.5%。特别是在今年春季，由于国内新冠肺炎疫情的影响，教育部“停课不停学”的号召，网上教学平台已成为“互联网+教育”的一个重要阵地。随着互联网技术的迅速发展，在线教育还存在着企业的盈利方式不对，用户的需求增多的问题，基于在线教育的现状状况，需要解决在线教育用户需求增多，企业的盈利方式错误，所以对如何利用大数据的优势来解决在线教育问题是值得深入发展的课题。

2 国内外研究现状

2.1 国内的研究现状

管佳，李奇涛认为我们应该积极地掌握网络教育的发展趋势，并结合学者的兴趣和喜好，为我国教育信息化建设提供参考，这对教育信息化来说是既是挑战更是机遇[8]。吕海燕等认为要引起学员对课程的重视，课程就应学员的兴趣爱好以及课程资源的自身特色[9]。张曦认为在线教育信息化的发展，同时也促进了中国教育事业快速发展[10]。郑勤华等认为由于在开学后的复课时间里，学校的课程安排相对较少，学生可以根据自己的兴趣爱好自主选择在线教学平台进行学习[11]。汤艳慧认为线上教育得以发展，应充分利用大数据的优势，对教学平台数据进行分析和处理，来发挥在线教育的应用价值[12]。苏泳睿，赵玲认为在线教育用户选择行为的研究呈现网络性特征，可以使用复杂网络的方法予以研究；在线教育用户选择行为是具有传播性的；关于在线教育用户选择行为的逻辑原初性，通过数据验证，可以认定逻辑起源于自身需求[13]。邬建中,秦倩认为在新教育传播技术网络支持下，以远程在线教育等形式，达成个人、 家庭、社会之间的协同学习，并形成以在线教育为媒体平台，集成各种应用场景，然后通过人与物的联结、物与物的联结完成个人教育、家庭教育、社会教育的融合与互动，探 索人工智能时代下的在线教育产业创新发展路径[14]。高振等人认为“新基建”和“双减”政策的施行为我国在线教育的健康发展做出了指导，希望未来政府、教育职能部门、机构、学校以及师生等主体，充分利用现代信息技术进行高效、协同处置治理，实现在线教育质量的稳定提升，促进线上线下教育融合创新，推进在线教育优质资源共享，最终实现教育公平[15]。

孙建平等人认为从教学质量管理机制、大数据组织框架、O2O 智慧教学平台入手创新在线教育模式，以期实现在线教育模式健康、稳健、可持续发展，为公民教育打造自由灵活的智能化教学方式[16]。袁婧怡等人认为在后疫情时代袁在线教育作为互联网与教育相结合的产物袁依然需要承担起链接安全与求知的责任[18]。邱宇,杜学元认为我们要加强对在线教育企业的监管，营造良好的在线教育生态，在线教育企业也要不断提升和完善自身，提高学生的学习效率[19]。刘文静等人认为线上教育和线下教育相结合的方式有可能是更符合现代年轻人的学习习惯的新的教学模式，线上教育的确能克服传统课堂教学的一些局限性[20]。肖娥芳,夏菲认为高校管理部门应不断加大对在线教育的重视程度，应进一步地提升教师群体现代教育技术水平，不断优化在线教育效果评价机制，以评促教，以评促学[21]。文博奚认为打通线上线下课程的数据隔离使得我们可以更精准地对学生的学习行为和老师的教学行为进行分析实现更高质量的教学活动[22]。闫鹏展认为在线教育网站可以支持学习者进行个性化的自主学习，满足随时随地学习的需要，但是由于在线教育网站黏性不够，导致学习者的学习动力、学习持续性下降，影响了在线教育的效果[23]。董坤景,赵丽娜认为在线教育行业在激烈的市场竞争中加速创新升级，内容和形式渐趋多样化。在线教育模式逐渐走向 OMO 模式：线上线下深度融合，实现产品和服务的标准化个性化，提高教学质量和运营效率，增加用户流量和粘性[24]。许梅,金本能认为在线教育打破了时空界限，满足了大家对于快速学习、个性化学习的需要，具有很强的生命力[25]。

2.2 国外的研究现状

Junjing Zhao（2020）基于共享经济背景下，认为在线教育充分利用互联网信息技术，创造了多种多样的教学模式充分，要充分发挥学生的学习兴趣，培养良好的学习习惯，促进学生的学业进步，促进了学生的学习和发展[26]。Anna Liu, Rui Zhang（2021）以猿辅导为研究对象，认为如果平台要实现长期稳定的发展，需要很好的满足需求，对于教育需求，只有通过提高产品质量和用户满意度体验，可以获得用户支持并产生用户价值；向潜在客户推荐产品，减少客户流失[27]。Oksana Pirogova 等（2021）在俄罗斯的网络教育基础下，认为在线教育市场已逐步扩大，且在线教育平台可以为学生提供全面便捷学习所需的功能，让用户在线上有个良好的体验[28]。Chen Lujun（2021）认为在大数据的背景下，针对不同的学生特点，制定相应的教学策略，让他们得以参加合适的培训，并能更好地理解他们的需求。通过实验研究，发现在大数据环境下，个性化网络教学平台有了显著的改进；从具体来看,在人工智能的背景下,线上用户数量增加了 9 %[29]。Wise

Tiffani M.,Opton Laura 认为在线教育的最大机会之一是教学中创造力和创新的开放平台。教师应与校园信息技术部门和教学设计师合作，了解增强在线教学的创新技术[30]。

3 评述

在线教育现在正是一片蓝海,随着社会的发展，针对细分场景的刚性需求产品的在线公司会越来越多，而也会带动软件、经济等的发展。在线教育企业正在蓬勃的发展，开放合作、相互借力会越来越多，不同的板块之间的结合，包括线上和线下结合，行业间的融合，场景间的融合等。所以在线教育未来的发展会更加的宽广。在线教育促进传统教学模式来变革，对于学习者，在线教育特有的时间空间优势能让学生随时随地自由学习。对于教授者，老师重建课程结构、改进教学自方式，老师上课不见得要讲很多，学生也可以把课程内容学会。所以网络在线教育会成为未来教育的指导趋势。人们可以接受更好的教育了不必深思，我们就能发现在线教育为什么能够在全球范围内迅速普及。即使是没有什么基础的初学者，也能有更多的选择来决定他们想要学习的内容。这些人不必再局限于传统的选择：在离家比较近的区域内选择。“现在的大学生都是互联网的’原住民’，他们对互联网有一种天然的亲近感，习惯和善于用’互联网+’的方式学习知识、捕捉信息”。

参考文献

[1] 刘盾. 在线教育的“烧钱争霸战”[N]. 中国教育报,2014-04-02(005).

[2] 朱新顺.“互联网+”时代在线教育研究与探索[J].现代信息科技,2019,3(22):146-147.

[3] 李准. 新东方在线科技控股公司 OMO 模式研究[D].吉林大学,2021.

[4] 蒋大成,王明宇.中国在线教育的现状和发展对策研究[J].电子商务,2015(09):68-69.

[5] 邵攸妮.互联网在现代教育的应用——在线教育的兴起与影响[J].中国多媒体与网络教学学报(上旬刊),2021(10):25-28.

[6] 邱宇,杜学元.试论在线教育企业面临的问题与对策[J].中国新通信,2021,23(22):161-163.

[7] 邢西深,李军.“互联网+”时代在线教育发展的新思路[J].中国电化教育,2021(05):57-62.

[8] 管佳,李奇涛.中国在线教育发展现状、趋势及经验借鉴[J].中国电化教育,2014(08):62-66.

[9] 吕海燕,周立军,张杰.大数据背景下教育数据挖掘在学生在线学习行为分析中的应用研究[J].计算技术与自动化,2017,36(01):136-140.

[10] 张曦.从在线教育临危上阵 谈谈教育信息化的“抗疫力”及发展建议[J].中国现代教育装备,2020(04):7-9.

[11] 郑勤华,秦婷,沈强,桂毅,周晓红,赵京波,王祎,曹一鸣.疫情期间在线教学实施现状、问题与对策建议[J].中国电化教育,2020(05):34-43.

[12] 汤艳慧.大数据在在线教学平台中的应用[J].办公自动化,2021,26(09):21-22.

[13] 苏泳睿,赵玲.复杂网络视域下在线教育用户选择行为特征研究[J].继续教育研究,2022(05):85-90.

[14] 邬建中,秦倩.人工智能时代在线教育产业的创新发展路径[J].传媒论坛,2022,5(04):53-55.

[15] 高振,娄方园,王书瑶,王娟.新基建背景下在线教育现状及治理策略研究[J].中国成人教育,2022(02):26-33.

[16] 孙建平,汪叙彤,郑晓云,肖琴.“互联网+”时代在线教育模式创新初探[J].山西广播电视大学学报,2021,26(04):36-40.

[17] 徐俊芳,董恒进.在线教育的发展及其质量评价研究[J].继续医学教育,2021,35(11):81-83.

[18] 袁婧怡,俞瞻宙,王静雯,常晶明,徐莹莹.后疫情时代大学生在线教育需求变迁研究——基于武汉市高校大学生的实证分析[J].湖北经济学院学报(人文社会科学版),2021,18(12):129-134.

[19] 邱宇,杜学元.试论在线教育企业面临的问题与对策[J].中国新通信,2021,23(22):161-163.

[20] 刘文静,杨婷,孙佳航,肖佳俊,王凯.远程在线教育的机会和挑战[J].科技风,2021(32):184-186.DOI:10.19392/j.cnki.1671-7341.202132060.

[21] 肖娥芳,夏菲.COVID-19 对高校在线教育发展的影响研究——以湖北工程学院为例[J].统计与咨询,2021(05):35-38.

[22] 文博奚.以学生为中心实现在线教育的高质量发展研究[J].湖南工业职业技术学院学报,2021,21(05):120-122+136.

[23] 闫鹏展,毕玉佩,谢昊,梁永慈.在线教育网站黏性提升策略研究[J].中国教育技术装备,2021(19):10-12.

[24] 董坤景,赵丽娜.后疫情时代河北在线教育市场现状分析[J].邯郸职业技术学院学报,2021,34(03):41-45+85.

[25] 许梅,金本能.疫情背景下我国在线教育发展面临的困难及其应对策略[J].安徽开放大学学报,2022(01):22-25.

[26] Junjing Zhao. Research on the Current Situation and Development Strategy of Internet Online Education Under the Background of Sharing Economy[P]. 2020 2nd Scientific Workshop on Advanced in Social Sciences, Arts & Humanities,2020.

[27] Anna Liu, Rui Zhang. Research on New Media Marketing Strategies of Ape Tutoring Online Education[J]. Scientific Journal of Economics and Management Research,2021,3(12)：131-136.

[28] Oksana Pirogova,Nataliia Temnova,Elena Markova. Online education in Russia: status and development trends[J]. E3S Web of Conferences,2021,258：10020.

[29] Chen Lujun. Application of Artificial Intelligence Technology in Personalized Online Teaching under the Background of Big Data[J]. Journal of Physics: Conference Series,2021,1744(4)：042208.

[30] Wise Tiffani M.,Opton Laura. Best practices in online education[J]. Nursing Made Incredibly Easy!,2022,20(3).

致谢

行文至此，也就意味着本科生的生活即将结束，始于 2018 年的秋天，终于 2022 年盛夏，在这大学四年中，承载着许多美好的回忆，度过了我人生当中最激情澎湃的四年。

首先，我想由衷的感谢我的论文指导老师曹永忠教授，从论文的选题、修改到最终的论文完成的每一个环节都离不开曹老师的悉心指导，他的渊博的专业知识给了我很多建议，让我的论文得以顺利的完成；虽然由于新冠肺炎疫情的缘由，只能在线上进行指导，但是曹老师以和蔼可亲的语气和尽职尽责的态度来指导和教导我们，值得我们尊敬；同时您的教诲也使我受益终身；所以我发至内心感谢曹老师的辛苦付出，承蒙教诲，不忘师恩；衷心祝愿曹老师身体健康，平安喜乐！

其次，我要感谢在我的大学生涯中每一个教授过我知识，帮助过我的老师，正因为有你们，我才能在广阔知识的海洋里遨游，感谢你们对我的专业知识的指导和建议，感谢你们对我的包容和教诲。由衷祝愿各位老师工作顺利、万事顺遂！

再者，我要感谢我的家人二十多年的养育之恩，感谢他们支持我的梦想，尊重我的决定，他们是我最坚强的后盾，每次经历过失败后，是你们让我有信心去坚持我自己想要的东西，让我有向前的动力，所以我一定会努力，成为你们的骄傲；感谢我的朋友三小只 120 度，和她们在一起的日子真的非常的快乐放松，感谢她们在我烦恼的时候给我的鼓励以及建议，在我伤心的时候无时无刻陪着我，在你们面前我能像一个小孩一样，无忧无虑表达自己的想法。祝愿我的家人们身体健康，快乐每一天；祝愿我的朋友们实现理想，找到自己的归属！

最后，我要感谢因缘分相聚于此，在大学四年 616 室友们、元气美少女们，以及闽台 1 班的全体同学们。和 616 的美眉们一起聊天到半夜，一起躲卫生查寝在宿舍不出声、一起玩 UNO，考试前一起背题复习，一起在体测时互相鼓励，感谢她们对我的坏脾气的包容、和她们在一起的时光总是那么的短暂，那么的难忘快乐；虽然我内向话不多，但是你们在我的身边，我得以安心。愿我们前程似锦，高处相见！

在未来的日子，我会全力以赴，不负青春年华，努力向前冲，感谢学校提供的平台，让我们能够施展我们自己的才华，以及认识到超赞的朋友，见证了我们的青春。最后，不忘初心，牢记使命，祝自己毕业快乐！