

本科生毕业设计

|  |
| --- |
| 面向MOOC的大数据分析 |

|  |  |
| --- | --- |
| 院 系 | 电子信息与通信学院 |
| 专业班级 | 电信1501 |
| 姓 名 | 刘未未 |
| 学 号 | U201513211 |
| 指导教师 | 刘澍 |

2019年06月04日

**学位论文原创性声明**

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

作者签名： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于 1、保密囗，在 年解密后适用本授权书

2、不保密囗 。

（请在以上相应方框内打“√”）

作者签名： 年 月 日

导师签名： 年 月 日

摘 要

MOOC（大规模在线开放课程，massive open online courses）截至2019年已经走过了7个年度，并掀起了一波网上自主学习的热潮，在MOOC逐步发展的过程中，使用MOOC平台的学生已经产生了大量的学习数据，对于教学工作者而言，这些数据无疑是非常有吸引力的。统计、整理、分析和挖掘这些数据等探究工作将帮助发现学习者学习特征及学习行为规律，并能够对改进教学设计和学生自适应学习提供更多的指导。

本文回顾了MOOC的发展历史，并对相关的国内外研究现状进行了梳理。本文选取一门发布在国内知名MOOC平台的课程，并对其单元测试行为，观看视频行为，讨论区域讨论行为等几个组成部分反映MOOC学习者的行为特征进行分析。 本文使用统计分析、聚类分析、灰色预测、多元线性回归分析等方法探索影响这些行为特点的诸多因素。并基于分析得出的结论，提出了对应的改进建议和策略。通过对数据的分析发现本文选取的课程具有关注度高，讨论区的总体参与度不是太高，视频和下载文档的行为是影响MOOC课程学习效果的重要因素，学期平均成绩稳定攀升等几个特点。

**关键词**：MOOC；学习行为；学习分析；多元线性回归分析；灰色预测

Abstract

MOOC (massive open online courses) As of 2019, MOOC has gone through seven years and set off a wave of online self-learning. In the process of MOOC's gradual development, students using MOOC platform have already A large amount of learning data is generated, which is undoubtedly very attractive to the teaching staff. Statistics, collation, analysis and mining of these data, this kind of inquiry work will help to discover learners' learning characteristics and learning behavior patterns. And can provide more guidance for improving instructional design and student adaptive learning.

This paper reviews the development history of MOOC and reviews the relevant domestic and international research status. This paper selects a post published on the well-known domestic MOOC platform.

The course, and its component test behavior, watch video behavior, discuss regional discussion behavior and other components reflect the behavior characteristics of MOOC learners. This paper uses graph analysis, multiple linear analysis, gray prediction and other methods to explore many factors that affect these behavioral characteristics. Based on the conclusions of the analysis, the corresponding improvement suggestions and strategies are proposed.

Through the analysis of the data, it is found that the courses selected in this paper have high attention, the overall participation of the discussion area is not too high, the behavior of video and downloading documents is the most important factor affecting the learning effect of MOOC courses, and the average grades of the semester are steadily rising Features.

**Keywords:** MOOC, learning behavior, learning analysis, multiple linear regression analysis, grey prediction

目 录

[摘 要 I](#_Toc10773857)

[Abstract II](#_Toc10773858)

[1 绪 论 1](#_Toc10773859)

[1.1 课题背景 1](#_Toc10773860)

[1.2 国内外研究现状 3](#_Toc10773861)

[1.3 研究目的、意义和主要内容 4](#_Toc10773862)

[1.4 论文结构【根据情况补充】 4](#_Toc10773863)

[2 MOOC大数据分析技术概述 6](#_Toc10773864)

[2.1 MOOC大数据分析框架 6](#_Toc10773865)

[2.2 大数据分析的方法 7](#_Toc10773866)

[2.3 大数据分析的工具 8](#_Toc10773867)

[2.4 基本方案制定 9](#_Toc10773868)

[2.5 本章小结 10](#_Toc10773869)

[3 方案实现与数据分析 11](#_Toc10773870)

[3.1 数据的可视化 11](#_Toc10773871)

[3.2 统计分析 15](#_Toc10773872)

[3.3 多元回归分析预测 20](#_Toc10773873)

[4 总结与展望 29](#_Toc10773874)

[4.1 研究结论 29](#_Toc10773875)

[4.2 针对MOOC平台的建议 **错误!未定义书签。**](#_Toc10773876)

[4.3 研究不足与展望 29](#_Toc10773877)

[参考文献 31](#_Toc10773878)

# 绪 论

本章首先介绍了信息时代的到来和MOOC的发展，然后分析了MOOC大数据的产生及研究意义，介绍了国内外对MOOC平台的相关研究工作，并对本文的主要研究内容及工作意义做了具体说明。

## 课题背景

### 教育大数据的爆发

今天的社会正处于信息爆炸的时代，每个领域都有大量的“大数据”[1]。麦肯锡说：“数据渗透到今天的每一个工业和商业的功能区[2]。人们挖掘海量数据并使用，表明新的生产力增长和消费者剩余浪潮的来临”。如何将这些数据转化为宝藏已经成为一种急需解决方案的问题。在教育领域，教育大数据越来越成为热门名词。这是从传统研究转变到科学实证研究的教育和科学研究领域研究的重要机会。与传统数据的宏观性质相比，大数据主要处于微观层面，这使得传统的经验式教学模式正在逐渐成为一种面向数据的转型[3]。

### 慕课的浪潮

随着时代的高速发展，以信息技术作为代表的技术革命浪潮正在迅速发展，它彻底改变了人类社会生活和生产的方式。在互联网这一时代，许多专家学者认为“在线教育将成为印刷以来最大的教育创新”[4]。现在，“大型在线课程”即MOOC已然成为最受欢迎的在线教育新形式之一。

2012年，MOOC喷发式涌现，广泛传播到世界各地的各大学，对全球高等教育的教学和学习产生了非常重要的影响，被盛誉为教育发展史上的“海啸”。随着Coursera，edX和Udicity被纷纷推出，知名大学也先后被引入，中国的许多知名大学也加进来为世界各地的学习者提供高质量的在线教育资源和服务[5]。

中国大学MOOC（慕课）是由网易和高等教育学会“爱课程网络”联合推出的大型开放式课程[6]。该课程学习平台于2014年5月启动，它与北京大学复旦大学、浙江大学、新加坡国立大学、微软亚洲研究院许多著名的大学的许多闻名中外的学院已经推出了数千个优秀的大学课程。让每一个有提升愿望的用户都能在此平台学习到中国最好的大学课程。其中，入驻高校和机构数：200+ （其中有120+高校和机构同时使用了学校云服务）；课程数量：1600+；课程视频数量：接近5万；报名选课人次：1200万+；注册用户数：500万。

### 选用课程概述

本课题选用了电子信息与通信学院在中国大学MOOC平台开设的《模拟电子技术基础》四个学期的后台数据作为研究内容。《模拟电子技术基础》课程是电气、电子信息类及部分非电类专业本科生在电子技术方面起入门性质的技术基础课，该课程具有很强的实践性和明显的工程技术特征。该课程学习分析和设计常用的电子设备，电路和系统，使学习者掌握电子技术的基本理论，基础知识和基本技能，并为电子技术的深入学习和实际应用打下坚实的基础。本课程2017年被教育部认定为首批“国家精品在线开放课程”。

我校的《模拟电子技术基础》课程是一门传统的优质课程，在2003年被评选为首批“国家精品课程”，2016年获得首批“国家级精品资源共享课”的称号。康华光教学团队编写的《电子技术基础》教材，曾先后获得国家优秀教材奖、特等奖、一等奖和国家科技进步二等奖四次国家级大奖，也是“九五”、“十五”和“十二五”国家级规划的教材，被国内众多高校采用。

本课程与康华光主编的《电子技术基础（模拟部分 第六版）》同步设计，课程团队秉承着重视教学内容与方法的研究及改革，重视学生工程实践能力培养的优秀传统，制作了内容精炼、质量上乘的教学视频，运用了大量动画生动地展现讲解的过程，直观明了，易于学习者理解。每章都有“本章小结”教学视频，帮助学习者串联整章内容，提高学习者的学习效率。同时，在多数章节的最后也有“习题课”教学视频，助力学习者举一反三，提升知识应用能力。另外，前6章配有PSpice和Multisim电路仿真软件的使用教学视频，以及电路仿真教学视频，便于学习者学习和应用。

## 国内外研究现状

### MOOC国内研究现状

笔者将中国期刊全文数据库（CNIKI）中的中国学术期刊网络出版总库作为本文数据来源，在2019年1月8日对摘要中含“MOOC”、关键词含有“数据”或“大数据”的CNKI的期刊论文进行检索，共检索到论文118篇。2013年发表的文章有3篇，2014年发表的文章7篇，2015年发表的文章16篇，2016年发表了文章31篇，2017年发表的文章为37篇，2018年发表文章19篇。可见，国内学者对运用大数据研究MOOC的热度经历了快速上升后有所会缓的过程。

2012年“学习分析”首次写入《地平线报告》，预示该技术会成为未来4至5年学习科学和在线教育的支柱，学习分析是在线教育分析学习数据的重要技术之一[7]。傅霖等学者发现影响MOOC学习效果的几个因素包括学习页面访问、视频观看、任务完成、讨论和学习成绩之间的相关性较小[8]。通过权衡每种学习行为的特征来发现影响几种学习效果的行为，然后基于这些行为特征进行聚类分析，将学习者分为不同种类的类别，并比较不同的类别学习者的行为差异[9]。候月姣等学者基于K-means聚类算法把学习者进行分类[10]。

### MOOC国外研究现状

在Specialsci进行检索，2015年有285篇，2016年有256篇，2017年有318篇，2018年有145篇。研究热度和国内情况相同在2018年有所下降。国外关于MOOC的研究主要集中在理论探讨层面。

J Jenny Mackness等研究人员通过电子邮件调用在线工具“CCK08”的MOOC课程以调查和总结MOOC重要特征：互动性、自主性、多样性和开放性[11]。也有人认为，在开发之初，MOOC课程的相关支持很少的情况是普遍存在的[12]。Mar Pez-Sanagustinl等学者分析到：MOOC浪潮的影响对传统教育产生了很大影响[13]。除了复制传统的教学方法，还可以利用一些社交工具（Facebook、Twitter和MentorMob）[14]。美国北亚利桑那大学设计的（GPS）工具系统，可用于评估课堂上的学生学习情况。教师通过该系统可以查到学业评级，学生出勤等详情给学习者提供指导。比利时鲁汶天主教大学的学生活动监视器系统提供了学生各种学习行为数据的可视化[15]。

## 研究目的、意义和主要内容

### 研究目的

通过收集MOOC平台后台数据并进行可视化，提取出记录该门课程的学习者单元测验、下载文档、观看视频、随堂讨论等学习数据，通过使用数据分析技术从而帮助决策者找到MOOC学习者各种学习行为的潜在因果关系并为MOOC的学习者，课程教师等相关人员提供建议。

### 研究意义

实现课程的评价并改进教学设计。MOOC的特点是学生规模大与种类繁多性，它对参与者的数量，时间和地点没有限制[16]。在传统的在线学习中，学习者群体的学习动机与学习者的心智以及知识水平大致相同;使用传统的教育评估方法研究MOOC学习者特征和学习效果的理论也不充分[17]。因此，本研究探索并量化了大量的学习行为数据，分析课程并提出课程设计的改进方法以供参考。

MOOC平台上收集了大量学习行为数据。这些海量的学习行为数据毫无疑问是了解学生的“关键”，通过分析学习者的行为数据找到学习效果与学习行为之间的关联[18]。

### 主要研究内容

依据本文的研究目的和研究方法，笔者注意到MOOC本身以及已掌握的数据特点，对中国大学MOOC平台上《模拟电子技术基础》的所有学习者产生的数据进行分类、整理、提取和统计分析。从学习者单元测验行为、观看视频行为、讨论区交互行为等方面研究MOOC学习者的行为特点，以及利用图表进行可视化分析、统计分析、多元线性回归分析、灰色预测等方法得出结论并提出相关的策略和改进意见。

## 论文结构

本论文由如下章节组成：

第一部分是绪论。介绍了本研究的研究背景，国内外研究现状，研究目的、意义和主要内容。

第二节为MOOC大数据分析技术。介绍了MOOC大数据分析框架、大数据分析技术与工具，并制定了研究方案。

第三节为本研究方案的实现与数据分析，对选定的课程学习数据进行可视化、统计分析、多元回归分析、灰色预测等。

第四节为全文的总结与展望，本章主要总结并反思了整个研究，反思本研究的局限性，并提出展望。

# MOOC大数据分析技术概述

MOOC的主体是学习者，学习者在学习过程中产生了海量的学习数据，同时学习者的学习背景和学习动机丰富多样。在MOOC平台运行过程中，学习者的学习行为都会被以多样的数据形式记录下来，因此基于MOOC大数据分析的研究是非常有必要的。本文先提出大数据分析的框架、分析方法、分析工具，然后将大数据分析应用到MOOC平台数据中来。

## MOOC大数据分析框架

大数据分析是指对规模巨大的数据进行分析。大数据可以概括为5个V，数据量大（Volume）、速度快（Velocity）、价值（Value）、真实性（Veracity）。大数据作为时下最火热的IT行业的词汇，随之而来的数据仓库、数据安全、数据分析、数据挖掘等等也成为了热点。

大数据已经不是简简单单是数据数量大，更重要的是对大数据的分析。只有通过分析才能获得很多更加深入、更有价值的信息。如今越来越多的应用涉及到大数据，大数据是互联网发展到现今阶段的一种表象或特征，在以云计算为代表的技术创新大幕的衬托下，这些很难被收集和使用的数据开始容易被利用起来了，通过各行各业的不断创新，大数据逐渐为人类创造更多的价值。这些大数据的属性，包括数量、速度、多样性等等都是呈现了大数据不断增长的复杂性，所以大数据的分析方法在大数据领域就显得尤为重要，甚至可以说是决定最终信息是否有价值的决定性因素。大数据分析的目标是对大量看似无关的数据进行集中的提炼分析，尽可能挖掘出数据之间潜在的关联。

面向MOOC的大数据分析框架主要可以划分为大数据来源、大数据类型、大数据分析以及大数据分析应用四个方面。MOOC大数据来源包括网页爬虫爬取、公开数据集、后台数据等部分。MOOC大数据类型丰富多样，包含学习者的多种学习行为数据比如观看视频、下载文档等行为数据。大数据分析的方法与工具也相对比较成熟，分析方法有可视化、聚类等。最后通过对MOOC大数据分析获得启示，并应用到MOOC平台改进中。面向MOOC的大数据分析框架如图 2.1所示。

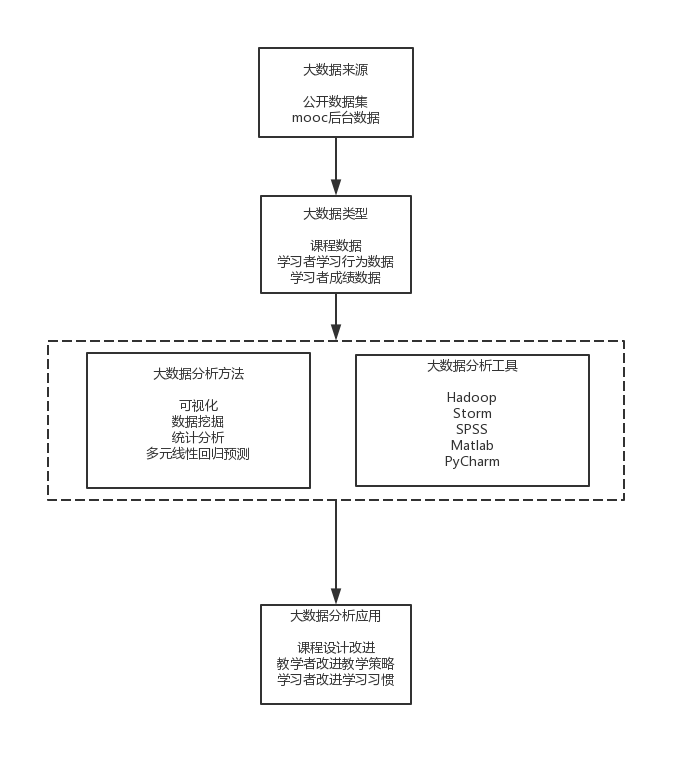


图 2.1 MOOC大数据分析框架

## 大数据分析的方法

（1）可视化

数据可视化是利用计算机图形学和图像处理技术，将数据转换成图形或者图像在屏幕上显示出来，并进行交互处理的理论、方法和技术。数据可视化的实质是借助图形化手段，清晰有效的传达与沟通信息，使通过数据表达的内容更加容易被理解。数据可以用散点图、折线图、直方图、柱状图、箱型图等来显示。本文以PyCharm为平台、python为实现语言，对MOOC平台产生的包括观看视频、下载文档、随堂讨论、单元测验等学习行为数据进行可视化，丰富的图表元素让数据“可观”。

（2）数据挖掘

数据挖掘是有组织有目的地收集数据，通过分析数据使之成为信息，从而从大量数据中寻找潜在规律以形成规则或知识的技术。通过对MOOC平台学习数据收集、整理、分析，抽取总结出有价值的信息，掌握学习者在学习过程中存在的优缺点，进而通过实施教学干预来提高学习者的学习效果，主要涉及的技术包括聚类、预测、关联规则挖掘等。

（3）统计分析

统计分析是指运用统计方法及与分析对象有关的知识，从定量与定性的结合上进行的研究活动。运用统计方法、定量与定性的结合是统计分析的重要特征。统计分析主要运用相关分析和回归分析，确定影响MOOC学习者学习效果的学习行为并建立模型，可以起到学习者参考自身学习行为的作用。

（4）多元线性回归分析预测

多元线性回归分析预测法，是指通过对两种或两个以上的自变量与一个因变量的相关分析，建立预测模型进行预测的方法。当自变量与因变量之间存在线性关系时，称为多元线性回归分析。分析影响MOOC学习成绩的学习行为并建立多元线性回归预测模型，并进行模型的检验。

## 大数据分析的工具

大数据作为时下最火热的IT行业的词汇，随之而来的数据仓库、数据安全、数据分析、数据挖掘等围绕大数据的商业价值的利用逐渐成为行业人士争相追捧的焦点。随着大数据时代的来临，大数据分析工具也越来越丰富多样。

（1）Hadoop

Hadoop是一个能够对大量数据进行分布式处理的软件框架。但是Hadoop是以一种可靠、高效、可伸缩的方式进行处理的。Hadoop是可靠的，因为它假设计算元素和存储会失败，因此它维护多个工作数据副本，确保能够针对失败的节点重新分布处理。Hadoop是高效的，因为它以并行的方式工作，通过并行处理加快处理速度。Hadoop还是可伸缩的，能够处理PB级数据。此外，Hadoop依赖于社区服务器，因此它的成本比较低，任何人都可以使用。

（2）Storm

Storm是自由的开源软件，一个分布式的、容错的实时计算系统。Storm可以处理庞大的数据流，用于处理Hadoop的批量数据。Storm很简单，支持许多种编程语言，使用起来非常有趣。Storm有许多应用领域：实时分析、在线机器学习、不停顿的计算、分布式RPC（远过程调用协议，一种通过网络从远程计算机程序上请求服务）、ETL（Extraction-Transformation-Loading的缩写，即数据抽取、转换和加载）等等。Storm的处理速度惊人：经测试，每个节点每秒钟可以处理100万个数据元组。Storm是可扩展、容错，很容易设置和操作。

（3）SPSS

SPSS可以用于分析运算、数据挖掘、数据预测的软件产品及相关服务的总称。SPSS功能很强大，能够提供众多的分析方法，比如多元回归、判别分析、聚类分析、Logistic回归等。

（4）MATLAB开发平台

MATLAB是由美国MathWorks公司生产的集成开发平台，特别是对数据处理方面，功能强大。应用于本研究的MATLAB主要实现了针对学生成绩的灰色预测模型，实现对学期平均成绩的预测，并对模型进行合理化检验。

（5）Python开发工具

Python在科学计算和数据可视化领域是一种强大的工具，本文调用numpy,pandas,matplotlib等第三方库实现数据可视化。

Pycharm是一款功能非常强大的集成开发环境，在本研究中将该平台作为Python编程语言的开发环境来使用。利用Python实现多元回归分析模型，分析与成绩相关的学习行为，实现对成绩的预测，并对模型进行检验，最后得出结论。

（6）Weka分析环境

Weka（即Waikato Environment for Knowledge Analysis，全名是怀卡托智能分析环境）是一种数据挖掘与分析的产品。Weka运行在Java环境下，是一个公开的数据挖掘工作平台，集合了大量能承担数据挖掘任务的机器学习算法，包括对数据进行预处理、分类、回归、聚类、关联规则以及在新的交互式界面上的可视化。

## 基本方案制定

MOOC后台导出的数据是不同学习数据类型的表格。以PyCharm作为开发平台，Python作为开发工具实现数据的可视化，根据图表进行分析得出结论并提出建议。通过对影响成绩的学习行为数据进行相关性分析，以聚类分析的角度为出发点，使用K-means聚类分析的方法尝试获得数据的聚类中心。使用多元线性回归分析对学习者个体成绩进行预测，并使用灰色预测模型对学期平均成绩进行预测。

## 本章小结

本章提出了一个MOOC大数据分析的框架，包括大数据来源、大数据类型、大数据分析以及大数据分析应用。对大数据分析方法、分析工具以及大数据分析方案进行概述。

# 方案实现与数据分析

前面论述了大数据分析的方法及分析工具，本章重点放在数据可视化，学习行为聚类分析、多元线性回归分析等方面。

## 数据的可视化

1. 成绩分布分析

箱形图最大的优点就是不受异常值的限制，能够准确稳定地描绘出数据的离散分布情况，同时也利于数据的清洗。箱形图是利用数据的五个统计量：最小值、第一四分位数、中位数、第三四分位数与最大值来描述数据的一种方法。由箱形图可以明显看出各学期成绩的中位数、上四分位数，都在逐渐增加说明成绩分布逐学期由低分段偏向高分段，学习者的学习效果也越来越好。另外2017年春季学期上限外有不少异常值，说明成绩过于集中在低分段，整体教学效果与学习效果不太理想，随着学期的进行，2017年秋季、2018年春季、2018年秋季成绩异常值变为无，说明学习者成绩分布更加合理，也体现教学团队教学方式与教学效果的改善。各学期整体成绩分布箱形图如图 3.1所示。

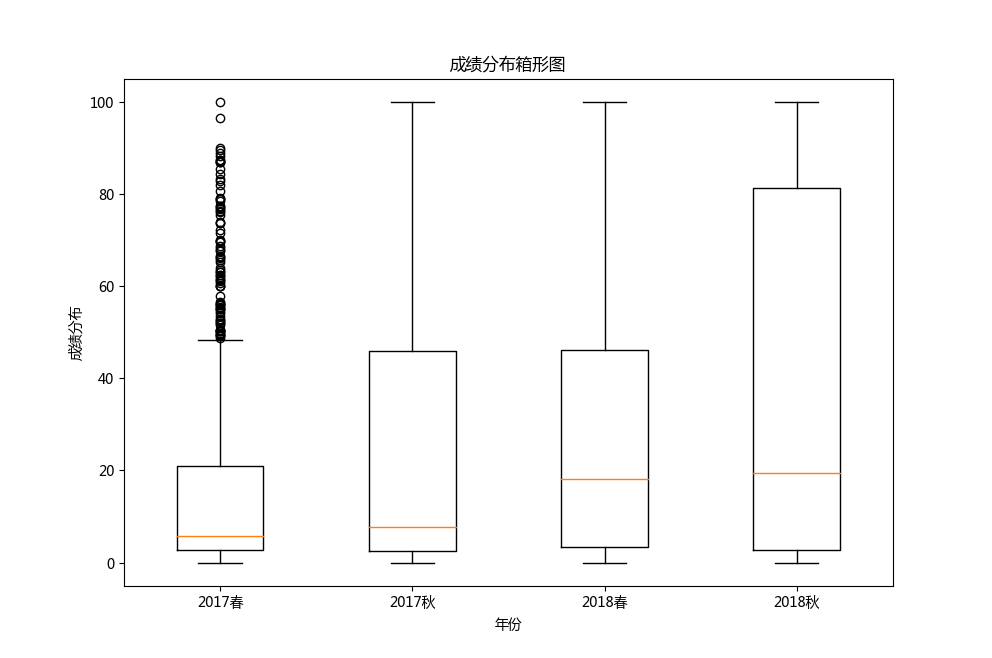


图 3.1 成绩分布箱形图

1. 单元测验分析

单元测验部分分为绪论测验题、运算放大测验题、二极管及其基本电路测验题、场效应管及其放大电路测验题、双极结型三极管及其放大电路测验题、放大电路响应测验题、模拟集成电路测验题、反馈放大电路测验题、功率放大电路测验题、信号处理与信号产生电路测验题、直流稳压电源测验题。以绪论测验题数据为例如图 3.2所示。

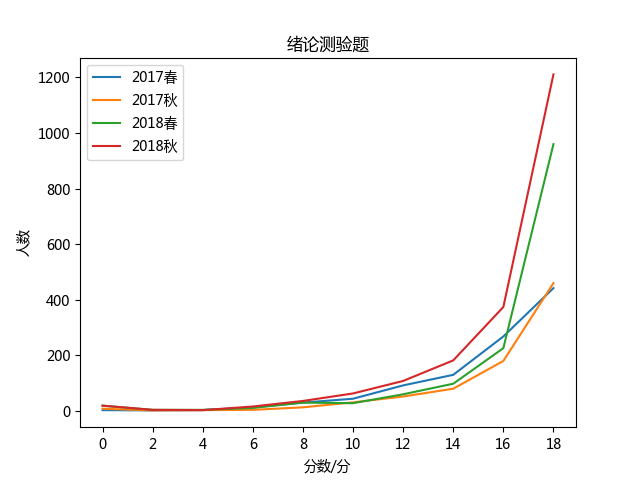


图 3.2 绪论测验题

通过整理四个学期的绪论测验题得到各学期在不同分数的人数，并将其进行可视化。绪论测验题一共有9道单选题，每道选择题两分，共计18分，如图所示为各学期在不同分数段的学生人数折线图。如果以总分18分的60%即10分及10分以上分数为及格分数，2017年春季学期、2017年秋季学期、2018年春季学期、2018年秋季学期各学期的及格率分别为95%，96%,95%，96%。四个学期的绪论测验题合格率都达到了95%及以上，说明学期刚开始学生们的学习热情比较高、学习效果也比较好。另外，各学期在不同分数段的学生人数均呈上升趋势，说明成绩分布是偏向高分数的。同时不同学期在相同分数上的学生人数总体上也呈现上升趋势，这与各学期参加学习的学生人数逐学期增加有很大的关系。另外，将四个学期相同的单元测验参加人数叠加并以柱状图的形式展示如图 3.3所示。

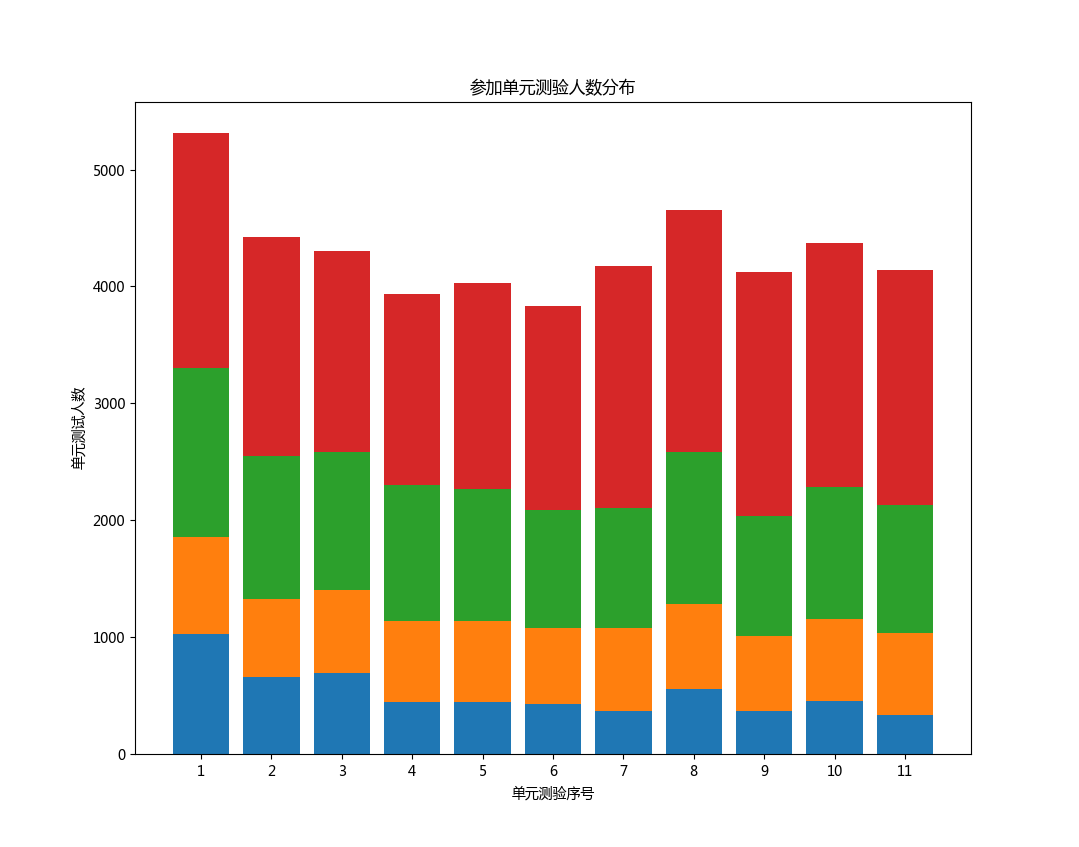


图 3.3 单元测验人数分布

由单元测验参加人数的柱状图可以发现，各单元测验参加总人数呈先下降后上升再下降的趋势，在第八个单元测验即反馈放大电路测验题参加人数形成峰值，说明学习者对此章节重视程度比较高，进一步说明该单元知识点是教学的重点，教学者和学习者都应以更高的重视程度对待该章节。

1. 选课情况

由MOOC后台数据各学期中第三个表格整理出各学期退选课情况。以2017年春季学期为例，截止课程结束，累计参加课程人数为15622人，累计退选人数为459人，最终实际参加课程人数为15163人。实际录入成绩人数为1689人，占总课程人数11.14%。平均每天选课人数为113人。2017年春季学期选课人数走势情况如图 3.4所示。

以日期为横轴，累计参选人数为左数轴，每天选课人数为右数轴，可以看出累计选课人数一直在增加，截止课程结束达到最大值15622人，选课人数在第一天达到峰值612人，接下来几天陡减，随后不断波动，最大值612，最小值27。，总体呈减少的趋势。

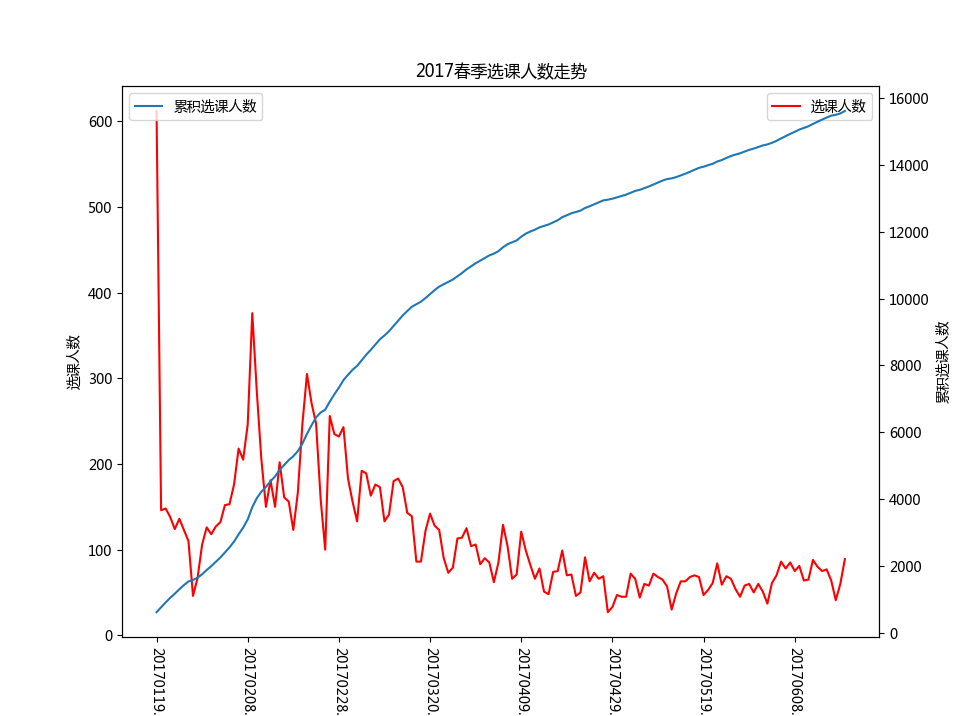


图 3.4 2017春季选课走势

1. 退课情况

类似地，以2017年春季退选人数情况为例，退选情况如图 3.5所示。可以看出累计退选人数一直增加，截止课程结束达到最大值459人。每天退选人数一直在波动，平均值为3人。2017年秋退课走势如图 3.5所示。

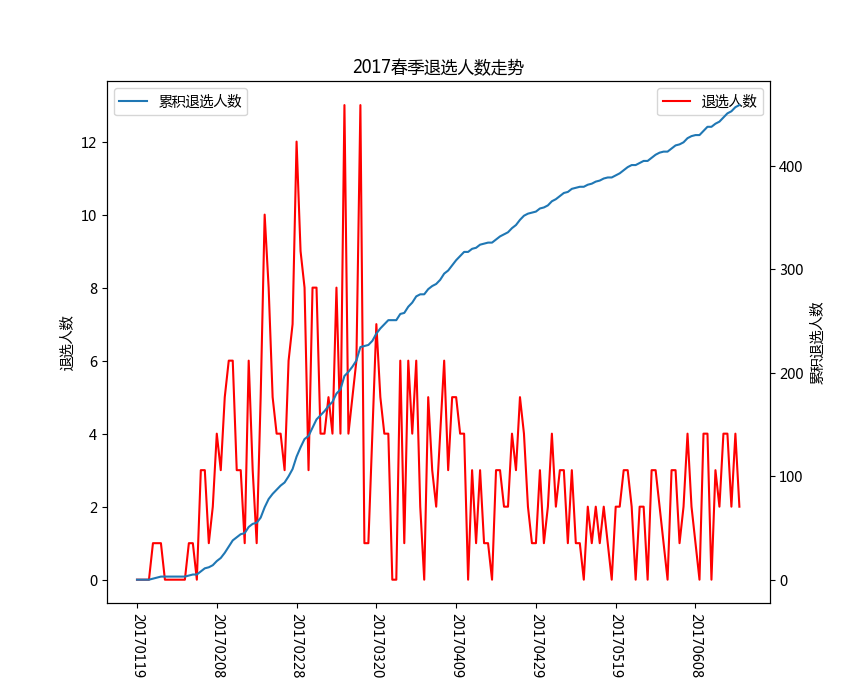


图 3.5 2017年秋退课走势

## 统计分析

### 基本的统计分析

（1）统计出四个学期的人数、成绩均值、成绩众数、成绩方差和标准差等数据，将四个学期的总体情况做个对比分析。

表 3‑1 总体数据统计

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | 2017年春 | 2017年秋 | 2018年春 | 2018年秋 |
| 人数 | | 1689 | 1500 | 2078 | 3398 |
| 均值 | | 14.01569 | 28.19525 | 31.59428 | 42.03917 |
| 众数 | | 3.64 | 3.23 | 2.83 | 3.64 |
| 标准差 | | 292.2524 | 999.7187 | 937.3789 | 1436.327 |
| 偏度 | | 17.09539 | 31.61833 | 30.61666 | 37.8989 |
| 峰度 | | 4.427709 | -0.74337 | -0.82079 | -1.73535 |
| 极小值 | | 0.14 | 0.13 | 0.14 | 0.13 |
| 极大值 | | 100 | 100 | 100 | 100 |
| 百分位数 | 25 | 2.83 | 2.38 | 4.455 | 3.33 |
| 50 | 6 | 10.035 | 22.1 | 33.33 |
| 75 | 21.34 | 48.28 | 51.615 | 83.1325 |

从描述统计结果表 3‑1可知，偏度都是正数的，由此能够判断学生每个学期的整体成绩分布都属于正偏态分布，即它们的峰值偏向左下方侧，并且每个学期的偏度值逐渐减小，因此偏斜度变得越来越不那么明显。 峰度值是正的变为负值，说明低分成绩聚集度在下降。

（2）各学期单元测验的合格率可以反映测试的难易程度，合格率越高的单元测验题目相对就更容易。各单元测验成绩的及格率如表 3‑2所示。

表 3‑2 单元测验合格率

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 参加测验学生合格率 | 2017年春 | 2017年秋 | 2018年春 | 2018年秋 |
| 绪论测验题 | 95% | 96% | 95% | 96% |
| 运算放大器测验题 | 89% | 94% | 92% | 90% |
| 参加测验学生合格率 | 2017年春 | 2017年秋 | 2018年春 | 2018年秋 |
| 二极管及其基本电路测验题 | 90% | 92% | 92% | 93% |
| 场效应管及其放大电路测验题 | 88% | 91% | 91% | 92% |
| 双极结型三极管及其放大电路测验题 | 90% | 93% | 91% | 90% |
| 放大电路频率响应测验题 | 87% | 90% | 86% | 86% |
| 模拟集成电路测验题 | 85% | 84% | 88% | 77% |
| 反馈放大电路测验题 | 81% | 78% | 79% | 75% |
| 功率放大电路测验题 | 86% | 83% | 85% | 79% |
| 信号处理与信号产生电路测验题 | 84% | 81% | 83% | 80% |
| 直流稳压电源测验题 | 80% | 77% | 79% | 75% |

合格率最高的单元测验均为绪论测验题，这可能与新学期学期热情较高以及题目难度较低有很大的关系。各学期合格率最低的单元测验都是直流稳压电源测验题，可能是学期末学习任务重，题目难度较大的原因。

（3）单元测验参加比例在一定程度上可以反映单元测验的重点程度，参加的比例越高的单元测验说明其对应的章节知识点越重要，大家想通过参加单元测验来考核下自己的知识掌握水平。

表 3‑3 单元测验参加比例

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 参加测验的比例 | 2017年春 | 2017年秋 | 2018年春 | 2018年秋 |
| 绪论测验题 | 58% | 54% | 66% | 57% |
| 运算放大器测验题 | 35% | 42% | 55% | 50% |
| 二极管及其基本电路测验题 | 37% | 43% | 53% | 48% |
| 场效应管及其放大电路测验题 | 23% | 42% | 51% | 44% |
| 双极结型三极管及其放大电路测验题 | 24% | 43% | 49% | 47% |
| 放大电路频率响应测验题 | 22% | 39% | 42% | 44% |
| 模拟集成电路测验题 | 19% | 40% | 44% | 47% |
| 反馈放大电路测验题 | 27% | 38% | 50% | 45% |
| 功率放大电路测验题 | 19% | 36% | 42% | 49% |
| 信号处理与信号产生电路测验题 | 23% | 38% | 45% | 48% |
| 直流稳压电源测验题 | 16% | 36% | 41% | 45% |

2017年春季、2017年秋季、2018年春季、2018年秋季各学期的有效成绩的人数分别为1689,、1500、2078、3398人。如图计算出各单元实际参加的比例。可知各学期各单元参加测验的比例整体呈下降趋势，这很大原因来源于学习者的学习热情逐渐下降，学习者应在学习课程中保持不懈怠的学习态度并且教学者在教学过程中适当提醒督促学习者学习和参加单元测验。

同时可以看出，各学期的测验成绩合格率呈先下降再上升再降低的趋势，并且合格率整体为下降趋势。说明中间一段时间及期末时间学生的及格率有待提高，进一步分析相关原因：随着课程的深入，学习难度增加，学习者的学习任务加重。另外，学习热情很大程度与外界刺激有关，在期初到期中学习压力不太重，学生们可能有所懈怠，临近期末，学习压力增加，学习热情被迫有所提升，但是持续一段时间，学生们的学习热情又可能有所下降。学生的学习热情随着时间的推移经历先下降再提升再有所下降的过程。

针对上述原因向教学者及学生提出三点建议：首先，针对单元测验合格率低于85%的对应教学视频进行调整，如增加教学实例、难点详解，使相应的知识点更加容易被学生理解吸收。其次，在整个教学过程在教学者承担的是一个教学质量把控者的角色，可以每两个单元测验提醒学生一次，尽可能让更多的学生参与每一个单元测验。最后学生要严格要求自己，在学习课程过程中不能懈怠，做到每一个单元测验都认真对待。

### 聚类分析

聚类算法是一种典型的无监督学习算法，该算法将样本划分为若干不相交的簇。聚类结果簇内相似度高，簇内相似度低。常用的相似度计算方法为欧氏距离。

从聚类分析出发，使用K-means聚类的方式尝试获得数据的聚类中心，以达到较好的聚类效果。

K-means首先随机选取k个质心，计算每个样本和k个质心的相似度（欧氏距离），选择相似度最高的质心所在的簇作为该样本的类别，形成k个簇。然后重新计算每个簇的质心（当前每个feature大小的平均值），再次调整每个样本所属类别，直到最大迭代次数或者调整的幅度小于阈值，算法停止。由于每次要计算所有样本与每个质心的相似度，故在大规模数据集上，K-means算法的收敛速度比较慢。K-means算法流程如图 3.6所示。

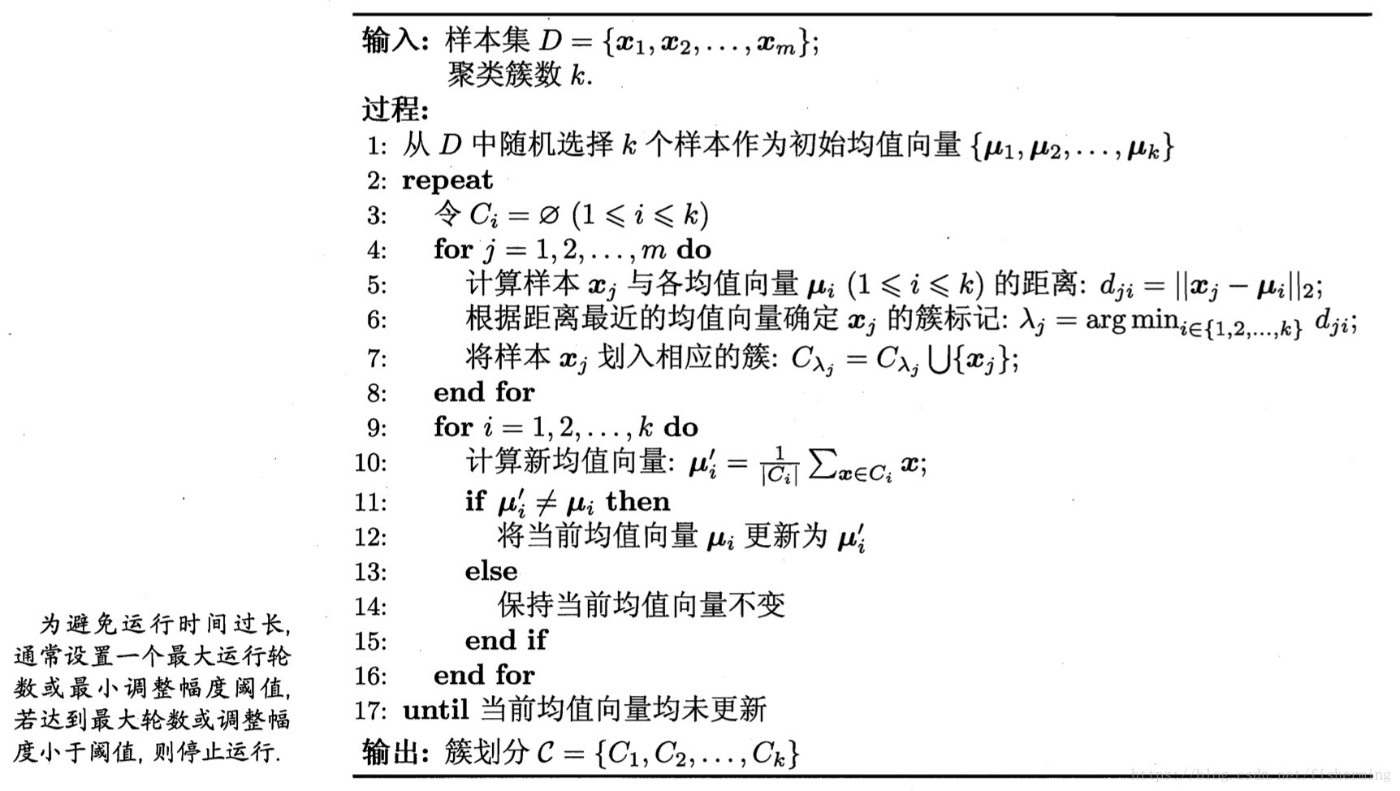


图 3.6 K-means算法流程

（1）以《模拟电子技术基础》课程2017年春季学期的学习数据为分析对象，发现与成绩相关变量有观看视频、下载文档、单元测验、随堂讨论等学习行为。将成绩、观看视频次数、下载文档次数、单元测验次数、随堂讨论次数进行K-means聚类，通过设置不同的K值得到不同的聚类结果如表 3‑4所示。簇内平方差和是衡量K-means聚类效果的量，簇内平方差和越小，聚类效果越好。当分类数从11到12时，簇内平方差和降低不明显并且迭代次数陡增，综合考虑确定K值为11，迭代次数为32，簇内平方差和为34.66。

表 3‑4 K-means聚类分析表

|  |  |  |  |
| --- | --- | --- | --- |
| 实验顺序 | 分类数 | 迭代次数 | 簇内平方差和 |
| 1 | 6 | 19 | 57.126947071229424 |
| 2 | 8 | 19 | 44.116541283895735 |
| 3 | 10 | 23 | 37.70725609925504 |
| 4 | 11 | 32 | 34.663096444731224 |
| 5 | 12 | 44 | 34.11072934711506 |

（2）K值为11时，各丛集的中心点如表 3‑5所示。观看视频次数、文档下载数与成绩呈现明显的正相关关系，说明这两者对成绩均有促进作用。另外，随堂讨论随成绩增加波动比较大，与成绩的相关关系不太明显，单元测验整体与成绩有一定的正相关。教学者应重视看待视频的拍摄与文档制作工作，保质保量。前三个丛集的人数占比接近七成，而且处于低分段，说明该学期学习者成绩总体不太理想。

表 3‑5 K-means聚类结果

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 丛集 | 观看视频数 | 文档下载数 | 单元测试 | 随堂讨论 | 成绩 | 数据量 | 占比 |
| 1 | 6.3644 | 5.1226 | 0.8161 | 0 | 3.0491 | 881 | 52% |
| 2 | 22.8082 | 19.137 | 6.6164 | 2.2603 | 13.1899 | 73 | 4% |
| 3 | 26.5333 | 21.8044 | 5.16 | 1 | 14.0584 | 225 | 13% |
| 4 | 32.2981 | 25.9615 | 4.2981 | 0 | 16.3342 | 104 | 6% |
| 5 | 28.8871 | 25.0161 | 13.1935 | 0.129 | 17.1729 | 62 | 4% |
| 6 | 48.1579 | 35.5526 | 4.3947 | 0.9474 | 28.7313 | 38 | 2% |
| 7 | 73.9083 | 56.1284 | 5.3394 | 1.2569 | 34.1061 | 109 | 6% |
| 8 | 76.6087 | 61.087 | 6.1884 | 0 | 37.4339 | 69 | 4% |
| 9 | 79.9091 | 55.4848 | 5.2727 | 0.9697 | 38.0973 | 33 | 2% |
| 10 | 70.8372 | 56.2326 | 5.6047 | 0.5814 | 38.7437 | 43 | 3% |
| 11 | 72.6538 | 57.0577 | 6.1923 | 0.5192 | 74.3712 | 52 | 3% |

1. 对数据进行基本统计，绘制出几个与成绩的散点图如图 3.7图 3.8所示。

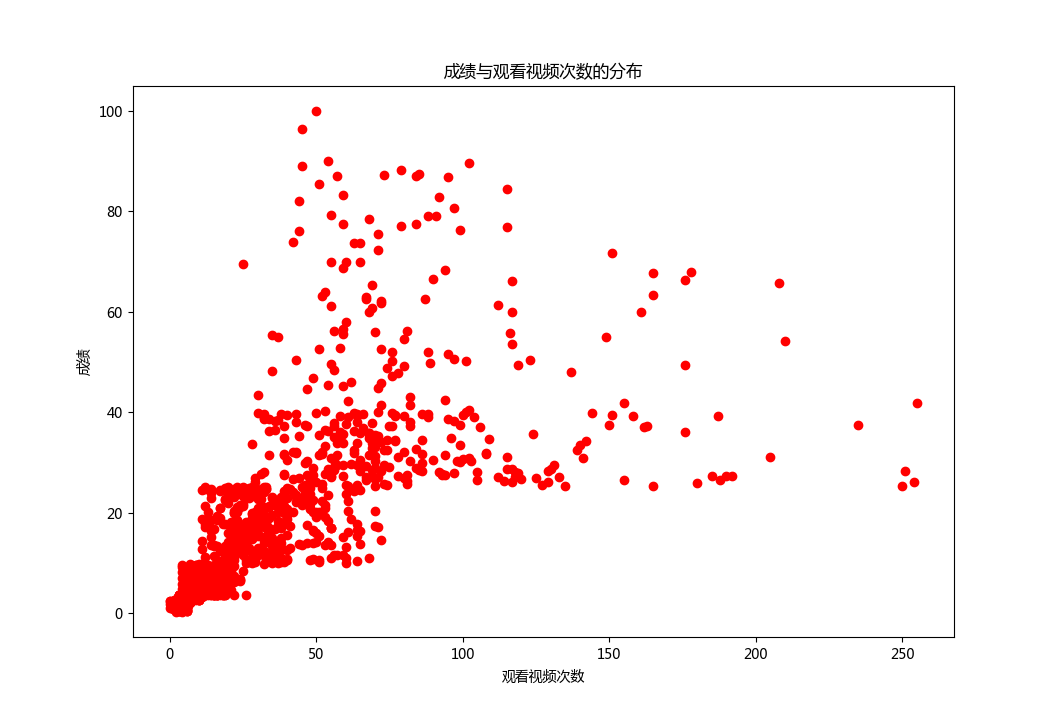


图 3.7 成绩与观看视频次数散点图

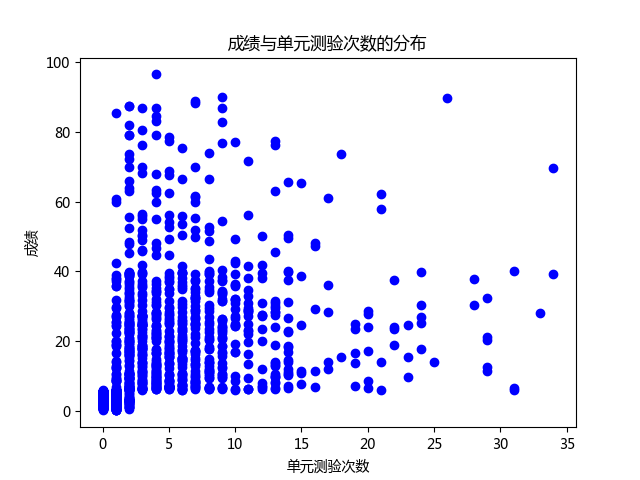


图 3.8 成绩与随堂讨论散点图

可以看出，图 3.7 成绩随着观看视频次数增加有上升的趋势，说明二者有很强的正相关性，观看视频次数最大值在250次左右，成绩在低分段比较聚集。图 2.1图 3.8 成绩与随堂讨论之间没有明显的线性，说明二者没有太大的相关性。

## 多元回归分析预测

多元回归分析是指在相关变量中将一个变量视为因变量，其它一个或多个变量视为自变量，建立多个变量之间线性或非线性数学模型数量关系式并利用样本数据进行分析的统计分析方法。

### 多元线性回归的计算模型

设y为因变量，x1,x2,…,xk为自变量，并且自变量与因变量之间为线性关系时，则多元线性回归模型为：



建立多元回归模型时，为了保证回归模型具有优良的解释能力和预测效果，应首先注意自变量的选择，其准则是：

1. 自变量对因变量必须是有显著影响，并且有一定的相关性。
2. 自变量与因变量之间的线性相关必须是真实的。
3. 自变量应有完整的统计数据，其预测值容易确定。

多元性回归模型的参数估计，可以用最小二乘法求解参数。以二元线性回归模型为例，求解回归参数的标准方程组为

 （1）

解方程（1）可得b0,b1,b2的数值。亦可用下列矩阵（2）求得

 （2）

### 模型的求解

（1）由上文分析可以发现，成绩与观看视频、单元测验、下载文档等行为有很大的相关性，将其作为成绩的因变量。

（2）以2017年春季学期学生的学习数据为研究对象，设观看视频、单元测验、下载文档分别为x1、x2、x3,成绩为y。得到回归方程



可以看出x1、x2、x3的回归系数均为正值，说明观看视频、单元测验、下载文档这三者的行为对成绩均为促进作用，符合常理。

模型的检验

1、拟合优度检验（R2检验）

拟合优度检验是测试回归方程与样本观察值的拟合程度。在所有解释变量和解释变量之间进行测试关联程度的一种方法。 测试方法是构造一个可以表征拟合的指标，该指标是通过分解总变差（总的离差）得到的16。







一个合适的回归模型，反映在平方和回归之和平方和的接近程度，则采用：



测试回归方程的拟合度。 如果所有样本观察都是位于回归方程，即：

此时，回归方程完全符合样本观察值，R2等于1：。

R2s残=6.874，s总=100.480

R2=0.93，接近1，说明该模型较好地拟合了样本观测值。

## 对平均成绩的灰色预测

### 概念

（1）白色系统、黑色系统和灰色系统

白色系统意味着系统中的所有功能都是已知的，即信息完全足够。

黑色系统意味着系统的内部信息完全不为外界所知。它只能通过与外界联系来观察和研究。

灰色系统中的一些信息是已知的，另一些是未知的。系统中的各种因素之间存在不确定的关系。  
 （2）灰色生成数列

灰色系统理论认为，尽管事物是客观和复杂的，但它们总是具有整体功能的，因此系统必包含一些内在规律。重要的是如何选择合适的方法来使用它。灰色系统通过整理原始数据（即灰色序列的生成）来寻求变化规律。 可以以某种方式生成所有灰色序列以削弱其随机性并显示其规律性。通用的数据生成方法包括累加生成、累减生成。

a、累加生成

把数列的各项（时刻）数据依次累加的过程称为累加生成过程（AGO）。由累加生成过程所得到的数列称为累计生成数列。设原始数列为,



原始数列x(0)的一次累加生成数列：





类似地有：



称为的r次累加生成数列。

b、累减生成

把对原始数据列依次做前后相邻的两个数据相减的过程称为累减生成过程IAGO。如果原始数据列为



的一次累减生成数列为：



注：从这里的记号可以看出，从原始数列得到新数列，再通过累减生成过程可以还原出原始数列。实际运用中，在数列的基础上预测出，再通过累减生成过程得到预测数列。

**3.3.2 灰色模型GM(1,1)**

**3.3.2.1 模型的建立**

灰色系统理论基于关联空间和平滑离散函数等概念定义了灰度导数和灰度微分方程，然后利用离散数据列以微分方程的形式构造动态模型。也就是说，灰色模型使用离散随机数来生成随机性显著减弱并且具有强大的规律的数字。并以微分方程的形式对其进行建模，以便于研究和描述该过程。

设为原始数列，其1次累加生成数列为

其中



定义的灰导数为

令为数列的邻值生成数列，即



于是定义GM（1，1）的灰微分方程模型为

,即或 (1)

在式（1）中，被称为灰导数，a被称为发展系数，被称为白化背景值，b被称为灰作用量。

将K=2,3，…,n代入（1）式有





……



引入矩阵向量记号

于是GM（1,1）模型可表示为****。

现在问题归结为求a,b的值。用一元线性回归，即最小二乘法来求它们的估计值为



注：实际上回归分析中求估计值是通过软件计算的，有标准的程序进行求解，如matlab等。

**3.3.2.2 GM（1,1）的白化型：**  
 对于GM（1,1）的灰微分方程（1），如果把灰导数的时刻,k=2,3,…,n视为连续变量t，则视为时间t函数，于是对应于导数，白化背景值对应于导数。于是GM（1,1）的灰微分方程相对应的白微分方程为

 （2）

**3.3.2.3 GM（1,1）预测的步骤**

（1）数据的检查和处理

为了确保GM（1,1）建模的可行性，有必要检查已知数据，检验处理。设原始数据为，计算数据列的级比



如果所有级比都在覆盖范围内，则为数据列可以建立GM（1,1）模型并且可以进行灰色预测。否则，需要对数据进行适当的转换处理，例如平移转换：使数据列，下降的比率可以覆盖。  
(2)建立GM（1,1）模型

不妨设满足上述要求，以它为原始数据列建立GM（1,1）模型



用回归分析求得a,b的估计值，于是对应的白化模型为



解为 (3)

于是得到



从而相应的得到预测值：



(3)检验预测值

1. 残差检验：计算相对残差



如果对所有的，则认为达到较高的要求：否则，若对所有的，则认为达到一般要求。

1. 级比偏差值检验：计算



如果对所有的，则认为达到较高的要求；否则，若对所有的，则认为达到一般要求。

**3.3.3 模型的分析和检验**  
 根据掌握的四个学期的统计数据，可以看出在正常情况下学期成绩的平均值较好的反映了相关情况，利用灰色理论建立GM（1，1）模型，由2017年春季、2017年秋季、2018年春季、2018年秋季成绩的平均值预测未来学期成绩的平均值。

**3.3.3.1 建立灰色预测模型GM（1,1）**

由已知数据，计算每个学期成绩的平均值，记为

 （3）

并要求级比

对作一次累加，则

 （4）

取的加权平均值，则，为确定参数，于是GM（1,1）的白化微分方程模型为

 （5）

其中a是发展灰度，b是内生控制灰度。

由于，取为灰导数，为背景值，则建立灰微分方程为



其矩阵形式为，

用最小二乘法求出参数的估计值：

（6）

则灰微分方程模型（4）的解为

则 （7）

由（7）式可以预测后几个学期的成绩平均值。

**3.3.3.2 模型的求解**

利用MATLAB实现灰色预测模型，代码见附录。并对模型的精度进行后验差检验。

2017年春，2017年秋，2018年春，2018年秋学期的平均成绩值

G(0)=[14.01569313,28.19525017,31.59428296,42.03916691]

可见级比λ（k）不在（0.6703,1.4918）并不符合条件，对数据进行平移变换，取C=50得X(0)=[64.01569313,78.19525017,81.59428296,92.03916691]，序列的级比（0.82，，0.96,0.88）符合条件。

进行累加求得：

X(1)=[ 64.0157，142.2109，223.8052，315.8444]

发展系数a=-0.0836,灰作用量b=68.4546

根据X(1)预测得到（0）数列[64.0157,76.9769,83.6870,90.9820]

所以预测的成绩为[14.0157,26.9769,33.6870,40.9820]

**3.3.3.3 模型的检验**

模型的检验是模型建立后所进行的不可或缺的工作，对灰色预测模型的检验一般有残差检验、后验差检验、级比偏差检验12等方法。

1、残差检验

残差残差在数理统计中是指实际观察值与估计值之间的差, 且具有误差的一些性质。利用残差所提供的信息，来考核模型假设的合理性及数据的可靠性称为残差分析。模型的残差判断标准有两种1、给定α，当残差平均值小于α成立时，认为模型为残差合格模型（一般α取0.01-0.05）2、如果模型的相对误差平均值在1%及以下，该模型的精度等级为一级；如果其大于1%且小于等于5%，该模型的精度等级为二级；如果其大于5%且小于等于10%，则为三级；当其大于10%且小于等于20%，则该模型精确等级为四级。

残差平均值e = 0.0457，对于判断标准1，取α = 0.05，e<α，改模型为残差合格模型。对于判断标准2，相对误差（0，1.56%，2.56%，1.15%），相对误差平均值为1.32%，该模型精度为二级。

2、后验差检验

X(0)数据平均值aver=78.9611

X(0)数据方差s12=100.480

残差方差s22=1.7433

预测精度等级：  
好 P>0.95 C<0.35   
合格 P>0.80 C<0.45   
勉强合格 P>0.70 C<0.50   
不合格 P<=0.70 C>=0.65

后验差比值：预测误差的方差同原始数据方差的比值C=0.0173。

小误差概率，P=1，预测精确等级为好。

1. 级比偏差检验

N=[ 0,-0.0714,-0.0836,-0.0773]，均达到较高的要求。

综上，灰色预测模型能较好地对学期平均成绩进行预测。

输入需要预测的学期m=1

得到f=[ 64.0157,76.9769,83.6870,90.9820,98.9129]

则预测的成绩为[ 14.0157,26.9769,33.6870,40.9820,48.9129]，图3-10为实际分数与预测分数图。

## 本章小结

# 总结与展望

## 研究结论

本文研究了MOOC大数据分析框架，总结了大数据分析的方法和工具。并以中国大学MOOC平台上的一门课程的后台数据进行实例分析，包括数据的可视化，统计分析、多元回归分析等。主要从学习者单元测验行为、看视频、下载文档、随堂讨论从这几个方面研究了MOOC学习者的学习特点。该研究得出以下结论：

1. 课程的关注度高，该课每学期都吸引了大量的课程注册者，但课程学习者的热情不长，随着课程的进行，学习者的学习热情越来越低，最终获得优异成绩的比例并不高，成绩主要集中在低分段，但是各学期学习者分数有上升趋势。
2. 观看视频、下载文档和单元测验等行为是影响MOOC课程学习者学习效果的重要因素，并且与成绩都是正相关的关系。讨论区总体参与度不高，参与者不是很积极，这可能与学习者的主动性不足有很大的关系，并且学习者成绩与随堂讨论情况没有很强的相关性。

## 针对MOOC平台的建议

1、从学习的角度来看，大多数学习者最初都抱着强烈好奇心和自我激励注册MOOC的，但经过一段时间后，很大一部分学习者失去了新鲜感和兴趣，甚至认为MOOC是他们的负担，所以在课程中间出现退出的情况，导致最终的整体效果并不理想。在这种情况下，可能是由于因为学习支持不足或缺乏对等支持和沟通。中国大学MOOC平台的每门课程都由学校的老师领导，老师带领一个小型教学支持团队，但仍有许多学生在学习过程中感到困惑和困难，问题得不到解决，这一部分很可能出现“破罐子破摔”。应该加强对这个问题的研究，支持确保教师或助教使用系统每天在线解决问题，并及时清除学习者的学习障碍。

2、在课程设计方面，增强讨论区的建设。讨论区是学生解惑的重要场景，本文研究的课程成绩与讨论区的相关性不高，说明讨论区的作用还没完全发挥出来，应加强讨论区整体的建设。例如教师团队就可以挖掘出来一些优秀的学习者帮助其他学习者并敦促他们学习，并增加讨论区的成绩占比，让更多的学习者参与讨论。如果一个困难的学习者得到及时的帮助，他就不会轻易辍学。另一方面，帮助其他学生的有能力的学习者可以获得奖励并认真落实，从而激励更多学习者更频繁地参与沟通中，从而使讨论区成为学习者之间，学习者与教学者之间互动的重要平台，发挥讨论区为学习者解惑的作用。

## 研究不足与展望

虽然本文较以往的研究，在某些方面有一些创新，但一些研究结果也存在许多不足，有待于更加进一步的探索和改进。

1、关于这篇文章的研究数据。本文数据基于本学院开设的《模拟电子技术基础》一课的后台数据，但数据本身存在局限性，有许多不完整的学习数据，个人学习信息较缺乏，这使得对产生学习数据的学习者分析不完整和细致。同时由于客观条件的限制，本研究只能获得一门课程的数据。本研究的结果可能并不代表整个MOOC领域的现状和问题，研究结果相对片面，因此将其在其他MOOC平台上推广可能有一些限制。

2、在选择学习数据分析模型时，由于以往研究中使用的MOOC学习数据分析模型很少，本文使用的模型可能也不是最成熟的。然而，理论模型是一个复杂的课题，作者可以在以后的研究中继续补充和改进。

从教育者的角度来看，有必要调整学生行为特征为教学策略的风向标。及时了解每个学生的特征非常重要。近年来，MOOC逐渐渗透到各个教学阶段，作为一种新的在线学习形式，这种教学模式具有很大的影响力。因此，必须尽快了解学习者的学习行为趋势。虽然本研究分析了学习者在课程中最基本的学习行为并得出一些结论，但由于数据本身的局限性，分析视角和模型预测还不够完善。如果可能的话，我会尝试从更多的学习行为数据和更多模型中预测，这将更有说服力。

参考文献

1. 武延军.大数据时代已经来来临——人机物融合的大数据时代[J].高科技与产业化,2013,05:46-49.
2. 焦建利.MOOC:大学的机遇与挑战[J].中国教育网络，2013,04:21-23
3. 张敏敏.国内MOOC到的研究现状分析[J].软件导刊（教育技术）,2015,10:51-52.
4. 王璐，杨洋，余佳每 大数据视角下的MOOC学习者行为特征分类研究[A],成都，2018.
5. 孙笑音，周围.于MOOC的大数据分析技术[J].计算机与现代化，2017年第四期，P89-95.
6. 张敏敏.国内MOOC的研究现状分析[J].软件导刊（教育技术），2015,10:51-52
7. 张强.MOOC学习者学习行为聚类分析[J].通化师范学院学报，2015,04,：37-39.
8. 傅霖，危妙，郭珊珊，黄敏祺.探析影响MOOC学习者学习成效因素[J].教育现代化，2015,06:60-62.
9. 王敏 基于行为日志数据地MOOC学习者学习行为分析研究
10. 候月姣.基于K-means聚类算法的远程学习者效果分析[D]。2011.1 北京邮电大学学报
11. Jenny Mackness,Sui Fai,John Mak,Roy Williams.The Ideals and Reality of Participating in a MOOC[M].2011.
12. Carlos Alario-Hoyos,Mar Perez-sanagustin，Carlos Delgado-Kloos,Hugo A Parada G.Analyzing the Impact of Built-In and External Social Tools in a MOOC on Educational Technlogies[M].The Invisible Librarian,2016,Pages 249-271.
13. The New Media Consortium.The Horizon Report 2010 Higher Education Edition[EB/OL].http://www.nmc.org/pdf/2010-Horizon-Report.pdf,2013-1-21.
14. Johnson L,Smith,R.willis,H.Levine A,Haywood K.The 2011 Horizon Report (2011)Austin,Texas:The New Media Consortium [EB/OL].http://www.nmc.org/publications/horizon-report-2011-higher-ed-edition.2011-11-20.
15. Leah P.M,Shane Dawson.Mining LMS data to develop an “early warning system”for educators A proof of concept[J].Computer&Education .2010,(54):588-599.
16. 候凤芝.在校大学生在线行为研究[D].浙江师范大学，2010
17. 高丹.大学生网络学习行为调查与研究[D].华中师范大学,2010.
18. 方旭.MOOC学习行为影响因素研究[J].开放教育研究，2015,03..
19. 刘京娟 多元线性回归模型检验方法 [A]湖南税务高等专科学校

**致谢**

当写到“致谢”两个字的时候，我很高兴我的毕业论文即将完稿了。但也意识到岁月匆匆的四年大学生涯要结束了，回首过往，万千感慨，对着四年中支持我、帮助我的人，心中充满了感激！无论是专业学习还是生活上，我所有的成长都和接下来要感谢的人密不可分。

首先，我要感谢我的导师刘澍老师，还记得半年前第一次见到老师的情景，历历在目，如今老师已成为我们的良师益友。感谢老师在这半年里以他独有的人格魅力潜移默化地影响着我。感谢老师在这半年期间每个星期都会开会指导论文工作。在论文的准备过程中，从文献阅读、开题答辩到数据分析、模型选取再到后期督促写文章，在整个过程中老师不辞辛苦帮我想点子、理思路，尽他所能帮我寻找解决问题的方法，多次询问研究进度并给予研究建议。可以说，没有刘老师对我论文倾注的大量心血，我的论文不会写的这么顺利，再次感谢我亲爱的刘老师！

其次，我要感谢电信学院所有的老师！在四年的大学生涯中，各位老师毫无保留、无私地传授专业知识给我们，他们教会我们的不仅仅是知识，更多的是教会我们如何学习。他们渊博的知识、开阔的视野、对学术孜孜不倦的态度为的是我们学习的榜样，在此致以我崇高的敬意和感谢！

我还要感谢我身边的朋友与同学，他们的陪伴点亮了我本来枯燥的大学生活，使我的生活变得丰富而温暖。感谢我的室友，在生活和学习上给我的帮助和鼓励，他们在平日的生活中给了我极大的帮助和宽容。感谢2015级电信学院同窗好友，在这个团结有爱的大家庭中，四年的共同生活让我倍感珍惜。

我深深地感谢我的家人！特别感谢家人支持我追求我喜欢的事情，感谢你们毫无保留的信任和支持，你们一直是我前进路上最大的动力和坚强的后盾！

我还要特别感谢我的一位朋友，她在我整个写论文的过程中给予了我莫大的鼓励与支持，是她的陪伴，让我在论文推进遇到困难而感到彷徨时选择坚守自己的初心，坚持以完成一篇高质量的论文为目标而不断努力。在此，我要特地感谢她一直以来的付出与相伴。

最后，还要感谢各位评审论文的专家和老师，对您们的工作表示衷心的感谢！