|  |  |
| --- | --- |
| 学号 | 2013302530129 |
| 密级 |  |

**武汉大学本科毕业论文**

**一种面向Mooc的个性化学习系统**

**的设计与实现**

|  |
| --- |
|  |

|  |  |
| --- | --- |
| 院（系）名 称 ： | 计算机学院 |
| 专 业 名 称 ： | 信息安全 |
| 学 生 姓 名 ： | 程秋平 |
| 指 导 教 师 ： | 何扬帆 副教授 |

二〇一七年五月

**郑 重 声 明**

本人呈交的学位论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确的方式标明。本学位论文的知识产权归属于培养单位。

本人签名： 日期：

摘 要

Mooc是发源于美国的一种新兴的在线学习平台。经过10几年的发展，包括中国在内的很多国家已相继开发了很多出色的Mooc平台。这些平台的产生在推动教育的创新和平等方面起到了很大的作用，同时能够很好地提高用户学习效率。随着Mooc的快速发展，各类课程资源不断累积，为用户建立学习计划变得越来越难。典型的问题包括：一、平台上各类资源的描述比较简单，用户在选择学习课程是缺乏足够的依据。二、课程学习的章节内容采用标准化的形式，每门课的章节顺序也统一不变。三、大部分平台没有考虑到不同用户的知识储备能力，也没有给不同用户对于同一门课想要掌握的程度的选择。四、课程测试部分仅提供答案的对错判断，并未提供错误知识点的再学习功能。本文针对上述问题提出了相应的解决办法，秉着个性化的原则，设计并实现了能够满足用户高效学习的Mooc平台。该平台的主要设计思想包括：一、支持个性化的学习路径定制；二、系统保存用户学习课程的记录和对课程的评分，使用协同过滤算法分析这些记录，挖掘用户潜在的兴趣课程；三、根据用户的背景和期望达到的学习目标为用户呈现课程的不同内容；四、使用Logistic Regresstion算法根据用户的学习目标和答案给出水平评级；五、每个测试题目附带相关知识点的关键词，平台能依据AC自动机匹配算法提供相关学习内容，便于学习者快速开始有针对性的复习。本文的研究结果能够为Mooc及其他学习平台的个性化学习过程定制提供借鉴。

关键词：Mooc；个性化学习；协同过滤；Logistic Regression

**ABSTRACT**

Mooc is a new online learning platform that originated in the United States.After 10 years of development, many countries including China have developed a lot of excellent Mooc platforms.The emergence of these platforms has played a significant role in promoting education’s innovation and equality,which is also a good way to improve users’ learning efficiency.With the rapid development of Mooc, the accumulation of various types of curriculum resources, the establishment of learning programs for users is becoming more and more difficult. Typical problems include: First, the platform’s description of various types of resources is relatively simple, the user in the choice of learning courses is a lack of sufficient basis. Second, the chapters of courses is standard, whose order and content is fixed. Third, most platforms do not consider the difference of basic knowledge among users, nor about giving choices of learning goal for users. Fourth,the part of test only provides the wrong or right judgement,and don’t provide the re-learning function of wrong knowledge. This paper provides a comprehensive solution to the above problem. adhering to the principle of individuality, this study implements a Mooc platform that supports personalized learning. The main design ideas of the platform include: First, the support of customizing personalized learning path; Second, the system saves the records of learning courses for each user, and analyzes these records using a collaborative filtering algorithm, through which the system can explore the user's potential interests; Third, showing different course content for user according to the user's background and expectations to achieve the learning goal .According to Fourth, the user's learning objectives and answers ,we will give a reliable estimation through Logistic Regression algorithm; fifth, each test question has relevant knowledge points, the platform can provide relevant learning content based on AC automatic matching algorithm , So that learners can focus on the review quickly. The results of this paper can provide a reference for the customization of Mooc and other learning platforms.

**Key Words**: Mooc; personalized learning ;collaborative filtering;Logistic Regression

目 录

[1 绪论 1](#_Toc481254630)

[1.1 研究背景 1](#_Toc481254631)

[1.2 国内外Mooc平台的发展 2](#_Toc481254632)

[1.3 个性化学习 3](#_Toc481254633)

[1.4 本文的研究内容 4](#_Toc481254634)

[1.5 本文的组织结构 5](#_Toc481254635)

[2 关键技术 6](#_Toc481254636)

[2.1 协同过滤算法 6](#_Toc481254637)

[2.2 Logistic Regression 8](#_Toc481254638)

[2.3 AC自动机匹配算法 10](#_Toc481254639)

[3 系统需求分析 11](#_Toc481254640)

[3.1 业务需求分析 11](#_Toc481254641)

[3.2 功能性需求分析 13](#_Toc481254642)

[3.2.1 用户业务模块用例图 13](#_Toc481254643)

[3.2.2 学习资源模块用例图 15](#_Toc481254644)

[3.2.3 系统UML类图 17](#_Toc481254645)

[4 系统设计 18](#_Toc481254646)

[4.1 开发环境与工具 18](#_Toc481254647)

[4.2 系统架构设计 18](#_Toc481254648)

[4.3 数据库设计 19](#_Toc481254649)

[4.3.1 系统E-R图 19](#_Toc481254650)

[4.3.2 系统数据库设计 20](#_Toc481254651)

[5 系统实现与测试 26](#_Toc481254652)

[5.1 各模块实现 26](#_Toc481254653)

[5.1.1 登录注册模块 26](#_Toc481254654)

[5.1.2 课程资源模块 27](#_Toc481254655)

[5.1.3 课程测试模块 31](#_Toc481254656)

[5.2 系统测试 33](#_Toc481254657)

[6 总结与展望 42](#_Toc481254658)

[6.1 工作总结 42](#_Toc481254659)

[6.2 研究展望 42](#_Toc481254660)

[参考文献 43](#_Toc481254661)

[致谢 46](#_Toc481254662)

# 绪论

## 研究背景

21世纪虚拟的信息技术成为了这个时代不可忽视的存在，人们的日常生活对网络的依附也日趋明显[1]。一方面，人们工作生活的各种需求正通过网络以更加便捷的方式得以实现，另一方面，人们获取信息也有了新的途径--通过可视化的操作直观迅速的获取各类信息。这种改变带来的影响不仅体现在人们的生活上，也潜移默化地把教育界带入了新时代[2]。人们开始习惯通过视频获得新的知识。这也导致了2012年MOOC[3]（Massive Open Online Course）的兴起[4]。

Mooc是一种以网络为依托的新型学习平台。它具有两大特性：一是开放分享，对学生身份没有严格限制，课程资源免费共享；二是可扩展性，抛开传统的学生和老师多对一的学习模式，Mooc的课程模式针对参与者是不定的。时代的发展引领着人们对于资源开放行为的支持，Mooc抓住这个契机在10多年期间快速成长[5]。国外的Mooc平台最早发起，著名的有Coursera、edX和Udacity三大巨头。国内很快也开始兴起很多Mooc项目，先有清华大学联合国内外其他高校成立“学堂在线”；后有上海交通大学自行创立“好大学在线”；此外还有很多包括网易云课堂、果壳网等高校以外的组织成立的Mooc平台。它们为学生提供了丰富的系统学习平台。

现今Mooc平台和网站越来越多，平台上的资源也在不断增多，它所带来的问题也愈见明显：一，Mooc平台上资源多样，对于学生来说快速准确地定位自己喜欢的课程尤为困难；二，教育具有多维性，用户的学习过程差异性很强，而当前的Mooc平台针对不同学生只能提供固定的学习资源和学习计划，因而难以对这种差异性作出有效应对；三，很多Mooc平台提供的测评模块只是简单地进行对错显示，没有根据学生对这门课程想到达到的学习程度进行综合性评估，学生对于自己的掌握程度不好准确地定位。由于这些问题的存在，“个性化学习”的研究在Mooc平台发展期间也逐步走向白热化[6]。

个性化学习需要根据学生兴趣、学习目标、学习能力等个体差异制定独一无二的学习策略，以达到提高学生学习效率和效果的目的[7]。因为不同的学习者在学习期间的习惯和方式等方面都会存在差异，所以，系统为其拟定的学习计划应该充分考虑每个学习者的特点[8]。另外，在学习某门课程期间，学习者可能因为自己状态的变化对学习方式作出相应的调整，比如要求更高，能投入的时间更少等等[9]。系统应当及时察觉这些变化，及时地对用户后续的学习计划作出调整[10]。

本文将针对基于Mooc的个性化学习展开研究，设计和实现一个具有个性化特征的学习平台，能够为不同特征用户提供独一无二的学习计划。该平台的主要设计思想包括：一、提供丰富的学习路径，避免用户学习盲目；二、系统把用户id、课程id、用户打分记录组成三维向量作为协同过滤算法[11]的输入，通过算法分析向用户推荐潜在感兴趣的课程；三、根据用户想到达到的不同的学习目的，对于同一门课程给出不同的章节内容；四、用户测评方面，使用Logistic Regression算法[7]，根据用户想要达到的目的和用户给出的测评答案，综合性的判断用户掌握程度是否合格。五、每个测试题目有自己的知识点关键词，使用AC自动机匹配算法[8]为题目匹配对应的课程，建立课程连接便于学习者快速复习。

## 国内外Mooc平台的发展

Mooc的出现不仅是教育界的一次改革，在商界更是一个契机，各大平台创始人都着力于打造更加优秀的Mooc平台[12]。2013年，Mooc很快蔓延到中国。5月，国内两大高校清华大学、北京大学对外宣布加入edX，同年10月，两所高校联合国内外其他著名大学成立了“学堂在线”，2014年4月上海交通大学自行研发上线了“好大学在线”[13]。近年，全球知名高校持续加盟MOOC这个在线教育大军。

1. Coursera

Coursera最初由斯坦福大学教授创办实现，后面陆续有全世界很多大学参与，包括中国的清华大学、北京大学等。Coursera的课程模式大概总结为：短时间（大概7-8分钟）的课程视频，用户学习完一段视频之后，会在中间穿插1~2个课程回顾性试题，主要题型是选择题,用户提交之后能获得及时的反馈。除此之外，Coursera还支持讨论互评的功能。最后Coursera也有课程测试的环节，也会得到及时的批改反馈[14]。毫无疑问的是，Coursera作为在线教育的巨头，其功能和学习资源方面是得到了用户的普遍认可，但是Coursera也有自己的一些缺点：比如在试题评估方面，只是简单地给出用户对错的显示，用户对于自己的掌握程度没有很好的定位，因为对于某些用户，并不要求自己掌握的十全十美，但是这样的用户对于Coursera给出的批改并不会对自己的掌握程度有很好的理解。

1. Edx

Edx创建于2012年1月，主办方是麻省理工学院和哈佛大学，他们力求建立一个全球影响力最大、最有知名度的在线教育非盈利组织[15]。Edx开发建设自己的课程发布与管理平台。其讲师主要是世界一流高校的教员，具有丰富的教学经历和一流的科研水平，能够开发出高效科学的在线学习课程。Edx的课程资源除视频资源之外，还包含其他形式的学习资源，比如视频、课后作业、讨论论坛以及wiki等。

1. Udacity

Udacity是由Sebastian Thrun和David Evans建立的学习社区，课程主要以数学和理工科科目为主，课程分初级、中级、高级三个等级。学习方式包含视频和非视频，非视频占的比例最大。Udacity的视频模块和练习模块紧密相连，这一点不同于Coursera和Edx，后两者这两个模块是完全不相干的。Udacity实现的功能主要有：教学信息广播、课程进度通知、学生能力测评、课后作业和课题讨论等[16]。

1. 果壳网

果壳网在2010年创立，该平台经过多年发展，研发出很多优秀的产品，主要包括MOOC学院、知性社区、研究生App等[17]。该网络站点具有开放和多元的特点。果壳网里面涵盖了包括Edx等平台在内的全部课程，并且克服了Edx课程全部是英文的缺点—果壳网把网站大部分视频都加了中文翻译，并且支持用户对课程打分、记录上课笔记等功能[18]。果壳网作为国内知名的在线学习网站，自上线开始就吸引了大量用户，但是经过研究发现，用户在果壳网上课程完成率非常低，部分用户给出的反馈是网站视频内容不够新颖，大部分用户的反馈是果壳网的教育并没有太大吸引力[12]。

1. 好大学在线

“好大学在线”自上线以来备受好评，该平台于2014年创立，由上海交通大学自行研发，是一个主要面向国内用户的在线学习系统[19]。平台包含的领域从文学到理工科不等，课程都是由国内顶尖大学的教师讲授。平台支持自主学习和学分学习两种学习方式，用户可以对课程打分，学习完之后也有对应课程练习。

## 个性化学习

个性化学习的英文含义是“personalized learning”。这是一种新的学习思维，与固定统一化的传统教育方式相比，这种学习方式更强调学习者的个性化特征。在这种个性化学习模式中，人们以自己现有的知识储备和自己所特有的本性特点为前提，以课程目的为导向，以自己能力和兴趣的自主为中心进行有效的学习行为[20]。

个性化学习包括传统线下学习方式的个性化和在线网络学习的个性化[21]。传统线下学习模式下，老师与学生主要以一对多的模式为主，在此环境下实现个性化学习的目的，老师需要记录多名学生的知识储备情况个学习能力，并能实时为每一个学生反馈学习情况，这对于老师的能力是个极大的挑战，到目前为止，个性化学习在这种学习方式下的实施尤为困难。

个性化也可以体现在在线网络学习的情况。基于网络的学习环境自身独有的特点也有部分个性化的体现：一、时间灵活性。在网络学习环境下，用户可以弹性地选择自己的学习时间[22]，相对传统学习环境，用户能够更加充分地利用学习时间；二、空间分布性[23]。网络学习环境的资源对于不同空间的用户是平等的，不同地理位置的用户均能对同一门课程进行学习；三、网络互联性。用户在网络学习环境下可以分享自己的学习心得、讨论课程问题，这有利于提高用户的自主学习能力；四、内容多样性，网络学习环境上的资源丰富多样，用户可以根据自己的学习需求找到目标课程。基于网络的个性化学习是在网络学习的基础上，根据用户学习的习惯和自身独特的特征，为用户提供更加具有个性的独特的学习计划。在这种学习模式下，系统需要记录用户的历史学习记录，包括视频记录、作业记录以及笔记记录等，初次之外，系统还要保存用户对课程的打分记录，通过对这些记录的分析，挖掘出用户的学习特征和知识储备情况，进而制定以用户为中心的个性化学习方案。

## 本文的研究内容

针对现有Mooc平台显现出来的特点以及个性化学习的目标，本文主要从以下几方面对Mooc平台进行改进：一、为了避免用户初次进入系统的盲目，系统会提供一些学习路径供用户选择，这些学习路径能够满足用户普遍学习目的。用户根据自己的兴趣和学习目的选择是否加入。二、依据协同过滤算法的原理，针对用户差异性，系统根据用户对学习资源的使用记录、打分情况等数据，建立用户学习资源的数据模型，通过距离运算公式计算出当前用户和其他用户的距离，使用一定的规则找出用户就近的n个邻居，从这些邻居中找出用户未学习过的学习资源推荐给用户。三、不同用户对于同一种课程有自己的学习目标，在学习一门课程之前，用户自己选择想要达到的学习目标，本系统包含了解、熟悉和精通三个学习目标供用户选择，用户选择的不同的学习目标，系统对于同一门课程针对用户选择的学习目标给出不同的章节显示。四、结合用户测评给出的答案情况，使用Logistic Regression算法可以评估出用户的学习能力，再结合用户选择的学习目标计算出用户解决对应难度的试题的概率，概率大于一定阈值就给用户“合格”的反馈，小于该阈值就反馈“不合格”。五、每个测试题目有自己的知识点关键词，每门课程的每个章节也有自己的关键词，使用AC自动机算法给题目关键词和章节关键词匹配，匹配到对应课程就让关键词链接到对应课程，便于学习者快速定位复习章节。

## 本文的组织结构

本文首先在第1章分析本课题在国内外的研究状况，依次介绍国内外有关Mooc平台的发展情况、个性化学习的概念以及解决个性化学习的方法，最后引出了本次针对个性化学习研究的内容。

接下来正式进入本次课题的详细介绍，在第2章详细阐述了本课题涉及到的关键算法以及原理。第3章是对系统的需求分析做了详细说明，对系统所要达到的目标效果结合图表进行了描述。第4章开始系统进入设计阶段，包含各个模块的功能和架构设计以及数据库的详细设计。第5章对系统的功能做了详细地展示，并且针对需求分析时的要求对系统进行了全面的测试。第6章是总结和展望，对本次课题做的工作进一步阐述和总结，并指出本次研究的缺点以及未来对本次研究的展望。

# 关键技术

本研究涉及到的关键技术有协同过滤算法、Logistic Regression算法、AC自动机匹配算法等。

## 协同过滤算法

1. 协同过滤算法原理简述。

协同过滤算法的工作原理非常简单，它的关键是充分使用大众智慧[24]，比如某个用户A想要找出自己喜欢的东西（比如电影）但是又毫无头绪，这时候就会想到找别人（用户B、C、D…）向自己推荐。一般情况下，A都会想要找和自己口味类似的人给自己推荐。

具体地说，协同过滤算法包含四步[25]：

1. 收集数据

数据指的是当前用户的在数据库存储的历史操作行为，比如之前看过的电影、给过的评分、收藏的项目等。特别要注意的一点，考虑到不同种类的数据粒度不同，使用算法时需要注意噪音的影响。

1. 计算距离

计算距离的方法主要有四种：

* 欧几里德距离：

|  |  |
| --- | --- |
|  | （2.1） |
|  | （2.2） |

欧几里德距离在数学上表示的是多维度两点的直线距离[26]，上面第一个公式中的x，y是n维向量，在协同过滤算法中x，y的每一维表示为计算用户距离的参数，x，y的距离表示用户的距离。公式2.1计算出来欧几里德距离，距离越小说明相似度越大，但是因为距离值的不可控制性，所以需要进行公式2.2的转换[27]。

* 皮尔逊相关系数：

|  |  |
| --- | --- |
|  | （2.3） |

从公式2.3可知，皮尔逊相关通过计算给出了线性相关程度，p的范围是[-1,1]，1表示x，y完全正相关，0表示x，y没有关系，-1表示x，y完全负相关[28]。在协同过滤算法中，x，y表示由两个用户或者两个item导出的向量，利用这个公式计算了两主体的相关性[29]。

* Cosine 相似度：

|  |  |
| --- | --- |
|  | （2.4） |

余弦相似度通过计算T值得出x，y的角度大小以达到计算x，y之间的距离的目的[30]。在协同过滤算法中，x，y分别对应两个主题对某个item的打分情况，通过计算打分向量的夹角，得到两个主题相似度[31]。

* Tanimoto 系数：

|  |  |
| --- | --- |
|  | （2.5） |

Tanimoto系数计算T值对应的是x，y向量的不相关性，该计算方案大多用于用于只用0、1两种值的情况[32]，在其他算法中使用效果不理想。

1. 找到相似物品或者用户

通过前面计算出了距离之后，就可以找出要推荐主体的邻居。可以使用两种方法[22]：一种是取和被推荐主体固定距离的邻居，距离在某个阈值之内的都算作邻居。第二种是取固定个数的邻居，即找出和被推荐主体距离最近的固定个数的邻居。

1. 进行推荐。
2. 本系统协同过滤算法使用

协同过滤算法的实现有很多现有框架可供使用，由于本次研究使用的语言是Java，所以系统使用了Mahout的Taste推荐引擎[33][34]，使用的是以用户为主题的推荐。

其实现原理是：把用户对所学习课程的使用记录、打分情况等数据，作为算法的输入数据，把用户对课程的打分当成参数计算用户之间的距离，根据距离找到与用户最接近的n个主体，找到n个相似主体之后，根据他们的打分数据，预测当前用户未学习的课程的打分情况，这样就生成了一个课程推荐列表。

Mahout的模型[35]如图2.1所示：

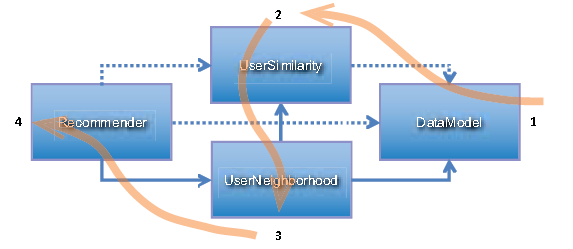


图 2.1 Mahout模型图

1. DataModel

用户对课程评分的抽象接口。本次研究中系统从数据表中导出部分字段值，这些字段包括当前主体的uid、课程id以及当前主体对课程的打分值，这些导出的字段形成了三维向量。

1. UserSimilarity

定义两个用户之间的相似度。计算相似度首先要计算两个主题之间的距离，本次研究系统通过把主题在系统中的学习记录和课程打分记录当做输入，使用欧几里得距离计算两个主题之间的距离。

1. UserNeighborhood

找邻居用户接口。查找邻居用户有两种方法实现，第一种是NearestNUserNeighobrhood，对每个用户找出固定N个邻居，这种方法在实现的时候其实设定了一个相似度阈值a，只有相似度大于a对应用户才被允许选入；另一种方法是ThresholdUserNeighborhood，对每个用户取一个对应的阈值，落在阈值范围内的就是该用户的邻居，系统采用的是第一种方法。

1. Recommender

Taste的推荐引擎的接口，扮演着核心组件的角色。在程序实现上，

把系统之前提供的DataModel作为输入，就可以得出对该用户的推荐。

## Logistic Regression

简称LR。LR的基本思想是基于线性回归[36]的方法进行回归和分类。分类主要是二分类。分类的原理总结为：根据已有的数据使用当前算法训练标准模型，根据这些模型对新样本进行划分[37]。逻辑回归能够很好的找到一个有足够好的区分度的决策边界，作出的分类也更加可靠。该算法的简单介绍如下：

* 1. sigmoid函数

它的公式：

|  |  |
| --- | --- |
|  | （2.6） |

它的图像[27]表示如图2.2所示：

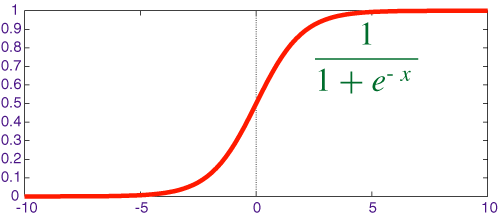


图 2.2 sigmoid函数曲线图

从上图看出sigmoid函数是个取值在[0,1]之间的s型曲线，在x=0是，函数值是0.5，往两边延伸函数值逐渐接近1或者0,。这是个很重要的性质。

* 1. 决策函数

一个机器学习的模型，实际上是把决策函数限定在某一组条件上，这组限定条件就决定了模型的假设空间，LR的模型的假设是：

|  |  |
| --- | --- |
|  | （2.7） |

可以看到，LR算法使用了上面的sigmoid函数，它的决策函数为：

|  |  |
| --- | --- |
| y∗=1,if P(y=1|x)>0.5 | （2.8） |

是预测函数的参数，x是算法的自变量，它们都是以向量的形式存在。在本次研究中，系统使用这个算法预测用户做题正确的概率，x是个一维向量，向量的值表示题目的难度，是二维向量，恒为1.0，于是就有：

|  |  |
| --- | --- |
|  | （2.9） |

* 1. 参数求解

求解的过程使用梯度下降法[39]：

|  |  |
| --- | --- |
|  | （2.10） |

求出偏导得到：

|  |  |
| --- | --- |
|  | （2.11） |

通过梯度下降法求解得到之后，根据用户选择对课程的掌握程度，系统给出不同的题目难度作为输入参数，即x，算出难度为x时用户求解这个题目的概率，概率大于一定阈值，系统就认为合格，否则不合格。

## AC自动机匹配算法

AC自动机是一个字符串匹配自动机，它的算法实现过程如下：

1. 使用关键词构造Trie树[40]
2. 在Trie树上建立失败指针，成为自动机。

构造失败指针的过程：假设当前要查找的节点的字母是x，让x沿着它的父亲节点的失败指针寻找，直到找到某个位置，它的儿子节点中含有x字母。则当前要查找的节点的失败指针就是那个字符也为x的儿子节点，另一种情况，假设一直寻找至root都没有找到符合要求的节点，则当前要查找的失败指针就是root[42]。需要注意的一点： root的子节点的失败指针都是root。

1. 根据AC自动机，搜索待处理的文本。

匹配过程：算法以根节点作为起点，根据扫描到的字符按照自动机依次搜索，如果扫描到的字符在自动机的某个分支里面返回为空，就按照失败路径递归搜索自动机，如果一直搜索到根节点，就忽略当前字符跳转到下一个。

# 系统需求分析

## 业务需求分析

在线个性学习系统主要包括三大块：一、用户业务模块，该模块包含注册、登录、学习路径、学习课程以及试题测试。二、学习资源模块，这个模块包含系统全部的学习路径、学习课程以及每门课程之后的试题检测。管理员负责上传和管理这些学习资源。三、学习资源推荐模块。该模块负责向用户推荐感兴趣的课程。其实现模型如图3.1所示：

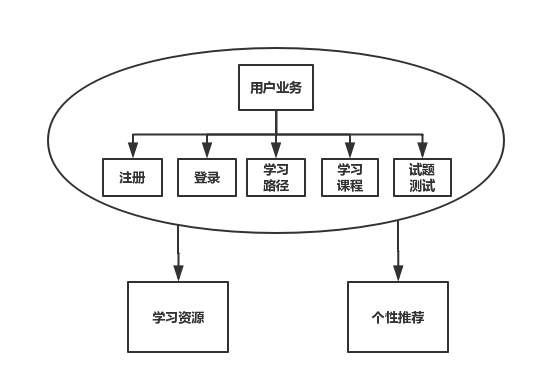


图 3.1 在线个性学习系统模型

1. 用户业务模块
2. 业务概述

这个模块是用户信息模块，包含用户注册、登录、用户在学习或者已经学习过的学习路径、用户在学习或者已经学习的课程、用户课程测试。在每门课程学习完后用户会被检测课程知识掌握程度。

（2）功能说明

注册：用户进入系统之前需要注册，该模块能够对输入的username的格式合法性进行检测，还能对两次设定的密码是否一致进行验证，注册成功用户就成为该系统一员。

登录：用户注册成功后,系统会根据数据库的记录对用户进行登录验证。

学习路径：这里主要是记录用户正在参与或者已经完成的学习路径，便于系统挖掘用户可能喜欢的课程。

学习课程：记录用户正在参与或者已经完成的课程，同时用户学习完当前课程所有视频之后会给当前课程以0~5分不等的打分。这一结果便于系统的个性化推荐。

试题测试：每门课程结束都会有课程内容相关的试题检测，用户提交试题后，测试系统一方面给出正误判断，另一方面使用Logistic Regression算法判断用户是否合格。

1. 学习资源模块
2. 业务概述

这个模块是系统的基本模块，包含系统全部的学习路径、学习课程以及每门课程之后的试题检测。管理员负责上传和管理这些学习资源。

1. 功能说明

学习路径：为了便于用户系统地学习某个领域的课程，系统为用户提供可选择的标准学习路径，路径下面包含若干主题，每个主题下面都有若干课程。

学习课程：课程主要有两个方面，一是“学习路径”下面的课程，另一方面是不依附于“学习路径”的课程。对于每一门课程，用户首先选择自己想要对这门课程知识的掌握程度（包括了解、熟悉、精通），根据用户不同的选择，系统会呈现不同的学习章节。用户学习完每门课程所有视频之后还可以对它做评分操作。

试题测试：每门课程结束都会有课程内容相关的试题检测，系统管理员负责课后quiz问题的协调。

3、个性推荐

1. 业务概述

这个子块也是本系统实现个性化的部分体现，其功能是显示用户可能想要加入的课程。

1. 功能说明

系统记录用户所有的加入课程行为、参与路径操作以及用户对所加入课程的打分情况，用户进入到该功能时，系统开始调用协同过滤算法，为用户分析潜在的感兴趣课程。

## 功能性需求分析

本系统包含的三个模块，其中用户业务模块和学习资源模块需要和用户交互，第三个模块是系统算法生成，不需要与用户交互，所以这里只有前两个模块的用例图和描述。

### 用户业务模块用例图

通过前面对系统的业务需求分析，这个模块的用户类型为普通用户，用户在这个模块与系统进行交互，包括加入学习路径的学习、选择课程学习，课程掌握程度的检测。其用例图如图3.2所示：

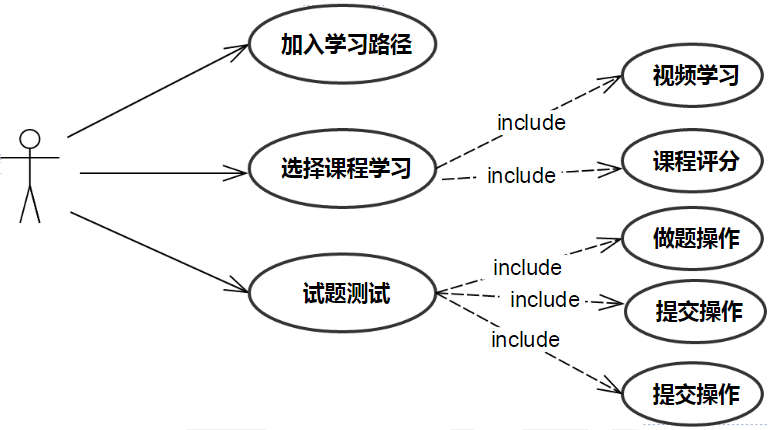


图 3.2 用户业务模块用例图

1. 加入学习路径用例，如表3.1所示：

表 3.1 加入学习路径用例

|  |  |
| --- | --- |
| 用例名称 | 加入学习路径 |
| 简要描述 | 用户进入页面，系统展示所有路径，用户自主选择是否加入学习 |
| 前置条件 | 用户必须正确登录，并且进入到这个页面 |
| 后置条件 | 用户选择路径的信息被保存到数据库 |
| 基本事件流 | （1）用户访问这个页面；  （2）如果用户之前加入过一些路径学习，这些路径显示“继续学习”按钮，如果没有加入过，显示“加入学习”按钮  （3）用户点击“继续学习”会跳转到显示学习主题页面，点击“加入学习”跳转到显示学习主题页面同时还会保存用户已加入该路径的信息 |
| 其他事件流 | 未点击“加入学习”或者“继续学习”按钮，用户随时可以点击“返回”按钮，信息不会作任何改变 |
| 异常事件流 | 当用户点击“加入学习”跳转页面有误或者数据库保存信息有误都会返回错误信息。 |

1. 选择课程学习用例，如表3.2所示：

表 3.2 选择课程学习用例

|  |  |
| --- | --- |
| 用例名称 | 课程学习 |
| 简要描述 | 用户进入页面，系统展示所有课程，用户自主选择是否选择学习 |
| 前置条件 | 用户必须正确登录，并且进入到这个页面 |
| 后置条件 | 用户选择学习这个课程信息会被保存，同时跳转到课程信息页面 |
| 基本事件流 | （1）用户访问这个页面；  （2）页面展示所有课程链接和简要信息；  （3）用户点击某个课程链接，跳转到课程详细信息页面，同时用户学习了这门课程的记录会被保存 |
| 其他事件流 | 未点击任何课程链接，用户可以点击“返回”按钮，信息不会作任何改变 |
| 异常事件流 | 当用户点击链接跳转页面有误或者数据库保存信息有误都会返回错误信息。 |

1. 试题测试用例，如表3.3所示：

表 3.3试题测试用例

|  |  |
| --- | --- |
| 用例名称 | 试题测试 |
| 简要描述 | 用户进入页面，系统显示该课程试题 |
| 前置条件 | 用户学习完这门课程 |
| 后置条件 | 用户提交试题答案，系统检测对错，把对错等信息保存到数据库，跳转到结果页面，每个试题下面都有对应keyword，每个keyword对应课程链接，便于用户复习。 |
| 基本事件流 | （1）用户访问这个页面；  （2）页面展示该门课程对应试题  （3）用户做题并且提交答案；  （4）系统调整到结果页面，显示对错结果以及是否合格，每个题目下面多有若干知识点的keyword，每个keyword对应课程链接，用户可以点击链接复习。 |
| 其他事件流 | 用户放弃做题，返回课程章节页面，信息不作修改 |
| 异常事件流 | 当用户答案提交失败、系统检查正误失败或者数据库保存数据失败都会返回错误信息，提示重新做题 |

### 学习资源模块用例图

通过前面对系统的业务需求分析，这个模块的用户类型是管理员，管理员负责课程视频上传、视频章节分类、生成学习路径、上传试题以及设置试题keyword等。其用例图如图3.2所示：

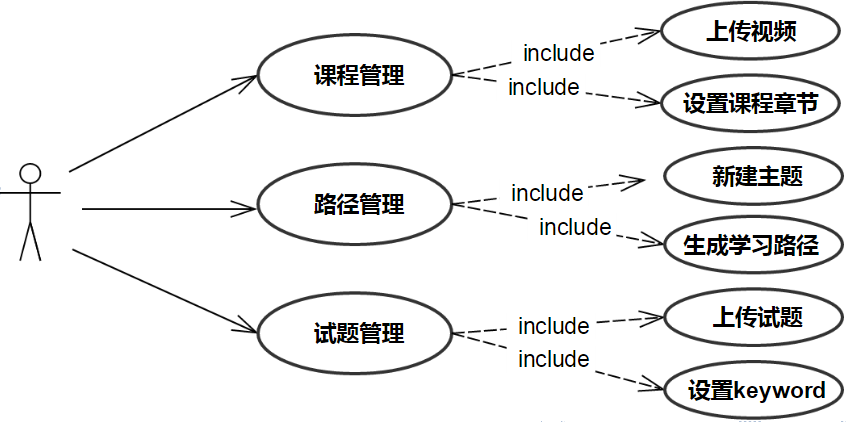


图 3.3学习资源模块用例图

* 1. 课程管理用例，如表3.4所示：

表 3.4课程管理用例

|  |  |
| --- | --- |
| 用例名称 | 课程管理 |
| 简要描述 | 用户进入页面，系统显示上传课程的页面，包含上传视频，输入视频名，以及设置课程章节，视频所属章节设置 |
| 前置条件 | 用户必须是管理员，管理员必须正确登陆 |
| 后置条件 | 信息输入完毕，存入数据库 |
| 基本事件流 | （1）用户访问这个页面；  （2）系统显示课程章节设置，用户输入章节内容，节奏上传视频，并给视频选择所属章节  （3）用户点击提交，数据库把信息存入数据库 |
| 其他事件流 | 用户放弃数据上传或者输入信息，点击“返回”按钮，信息不作任何改变 |
| 异常事件流 | 当信息存储失败时，返回错误信息并且提示重新输入或者上传 |

* 1. 课程管理用例，如表3.5所示：

表 3.5课程管理用例

|  |  |
| --- | --- |
| 用例名称 | 路径管理 |
| 简要描述 | 用户进入页面，系统显示新建主题、利用已有主题设置学习路径，并且选择已有课程导对应主题 |
| 前置条件 | 用户必须是管理员，管理员必须正确登陆 |
| 后置条件 | 信息输入完毕，存入数据库 |
| 基本事件流 | （1）用户访问这个页面；  （2）系统新建主题页面，用户新建主题信息，并且把已有课程导入到对应主题，利用已有主题新建学习路径  （3）用户点击提交，数据库把信息存入数据库 |
| 其他事件流 | 用户放弃数据输入，点击“返回”按钮，信息不作任何改变 |
| 异常事件流 | 当信息存储失败时，返回错误信息并且提示重新输入或者上传 |

1. 试题管理用例，如表3.6所示：

表 3.6课程管理用例

|  |  |
| --- | --- |
| 用例名称 | 试题管理 |
| 简要描述 | 用户进入页面，系统显示上传试题信息以及为试题设置keyword等信息 |
| 前置条件 | 用户必须是管理员，管理员必须正确登陆 |
| 后置条件 | 信息输入完毕，存入数据库 |
| 基本事件流 | （1）用户访问这个页面；  （2）用户先输入试题信息，包括试题内容和答案，以及试题包含的知识点keyword  （3）用户点击提交，数据库把信息存入数据库 |
| 其他事件流 | 用户放弃数据输入，点击“返回”按钮，信息不作任何改变 |
| 异常事件流 | 当信息存储失败时，返回错误信息并且提示重新输入或者上传 |

### 系统UML类图

整个系统主要实体类包括用户类User、视频类video、课程类Course、老师类Teacher、问题类Question、学科主题类Subject以及路径类path，他们之间的关系如图3.4所示：

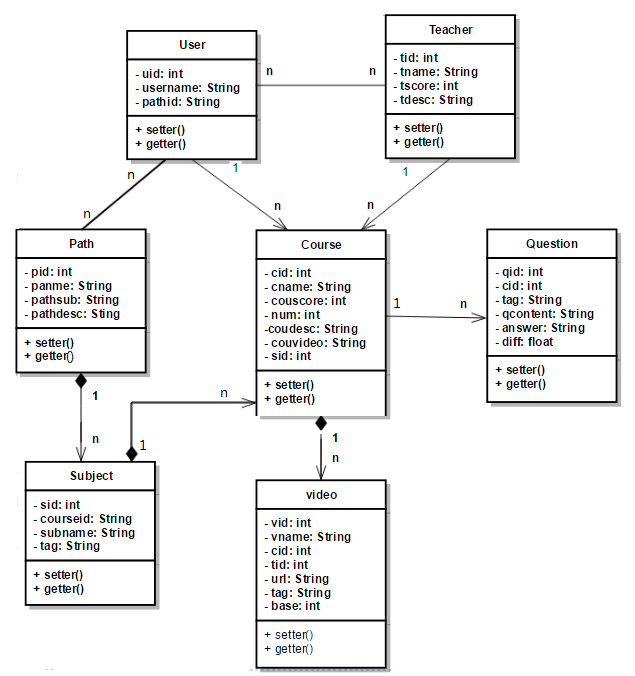


图 3.4 系统UML类图

# 系统设计

## 开发环境与工具

本次研究中，系统实现的平台是windows 8.0+MyEclipse2014。代码实现方面，使用的模型是servlet+jsp+JavaBean的MVC模型，系统前端框架使用的是AdminLTE-2.3.11；服务器端使用的是Tomcat7.0；数据存储方面，使用的是MySQL5.0+Navicat，通过hibernate 4.1.4框架访问数据库。

## 系统架构设计

根据上一章需求分析，整个系统可以分为三个功能模块：注册登录、课程资源、试题测试。其中课程资源这一块包括用户路径选择和对具体课程视频的选择。系统结构绘制成图，如图4.1所示：

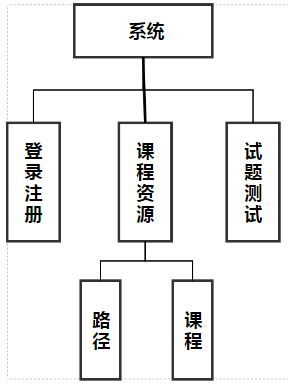


图 4.1 系统结构图

1. 登录注册模块，主要包括注册、登录、密码重置功能。如图4.2所示

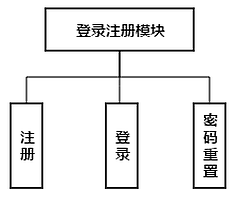


图 4.2 登录注册模块图

1. 课程资源模块是系统的基本模块，包含一些标准学习路径、从属于某些路径或者不从属于任何路径的课程，用户对一些课程的学习记录会被记录到数据库，系统会使用算法根据这些记录为用户推荐可能感兴趣的课程，用户学习完一门课程也可以为课程评分。结构图如图4.3所示：

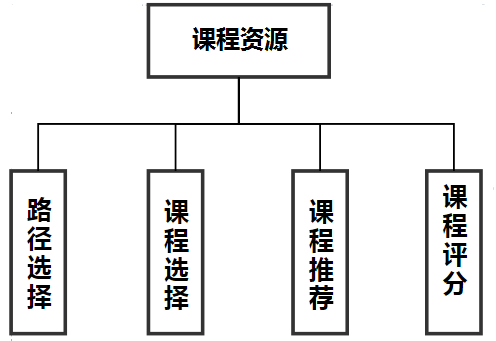


图 4.3 课程资源模块图

1. 试题测试模块是用户学完一门课程对课程的检测模块，主要包括做题、评估以及根据keyword复习三个部分，结构如图4.4所示：

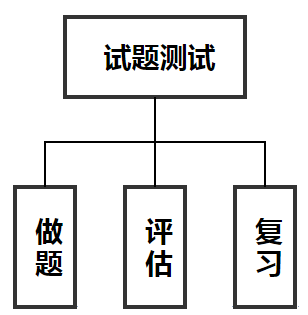


图 4.4 试题测试模块图

## 数据库设计

### 系统E-R图

根据上一章的分析，系统会有7个实体，分别是用户user、老师teacher、学习路径path、视频video、课程course、课程主题subject以及问题question。他们的属性以及他们之间的关系用E-R图表示4.5如下：

### 系统数据库设计

1. user表。

保存了用户相关信息，主键为uid，另外还包含用户名，pathid等信息。其中pathid为用户参加的路径id，是text类型，每个id之间用逗号分隔。结构如表4.1所示：

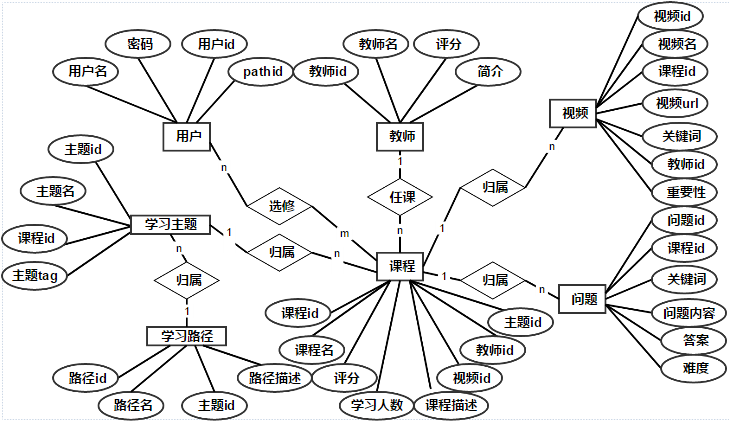


图 4.5 系统E-R图

表 4.1 user表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| uid | int(11) | NO | PRI | NULL | 用户的id |
| username | varchar(20) | YES |  | NULL | 用户名 |
| Password | varchar(50) | YES |  | NULL | 密码 |
| pathid | text | YES |  | NULL | 该用户选择的路径id，每个id之间用逗号分隔 |

1. teacher表。

保存了教师的信息。其表结构如表4.2所示：

表 4.2 teacher表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| tid | int(11) | NO | PRI | NULL | 教师的id |
| tname | varchar(20) | YES |  | NULL | 教师名 |
| tscore | float(2,1) | YES |  | 5 | 教师评分，最高分是5分，默认是5分 |
| tdesc | Varchar(255) | YES |  | NULL | 教师的自我介绍 |

1. 课程表。

描述了课程信息，每门课含有一定的章节，每一节内容对应一个视频。它的表结构如表4.3所示：

表 4.3 course表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| cid | int(11) | NO | PRI | NULL | course的id |
| cname | varchar(100) | YES |  | NULL | course的name |
| couscore | float(2,1) | YES |  | 5 | 该course的评分 |
| num | int(11) | YES |  | 0 | 学习该course的人数 |
| tid | int(11) | YES | MUL | NULL | 授课老师的id |
| coudesc | text | YES |  | NULL | 该课程的简要介绍(description) |
| sid | int(11) | YES | MUL | NULL | 该课程对应的主题（Subject）的id |
| couvideo | text | YES |  | NULL | 该course含有的video的id（每章之间用”|”分隔，每节之间用逗号分隔） |

1. 学习主题subject表。

保存的是每门课程对应的主题的记录，表中个字段如表4.4所示：

表 4.4 subject表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| sid | Int(11) | NO | PRI | NULL | subject的id |
| courseid | text | YES |  | NULL | Subject对应的course的id（用逗号分隔） |
| subname | varchar(100) | YES |  | NULL | subject的name |
| tag | text | YES |  | NULL | 该subject具有的标签 |

1. 学习路径表。

描述了学习路径的信息，结构如表4.5所示：

表 4.5 path表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| pid | int(11) | NO | PRI | NULL | 路径的id |
| pname | varchar(100) | YES |  | NULL | 路径名字 |
| pathsub | text | YES |  | NULL | 路径对应的主题的id，每个阶段之间用”|”分隔，阶段内的主题之间用逗号分隔 |
| pathdesc | text | YES |  | NULL | 路径的描述 |

1. 问题表。

每门课程之后对应有一定数量的问题，目前系统只涉及选择题，其表结构如表4.6所示：

表 4.6 question表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| qid | int(11) | NO | PRI | NULL | 问题的id |
| cid | Int(11) | YES | MUL | NULL | 问题对应的课程id |
| tag | text | YES |  | NULL | 问题所含知识点的关键词，关键词之间用逗号分隔 |
| qcontent | text | YES |  | NULL | 问题的内容，形式为问题|选项A|选项B|选项C…. |
| answer | text | YES |  | NULL | 答案，因为可能有多选，所以考虑用text，每个选项之间用逗号分隔 |
| diff | float(2,1) | NO |  | NULL | 问题的难度，最大难度是5.0 |

1. 视频表。

描述了每个视频信息，其中base是表示这个video在对应课程的重要程度，如果是0或者2说明这个video是可选择性的学习的，如果是1说明这个视频很重要，任何用户都要学习。如果用户选择对这门课是“了解”的程度，那么0或者2的视频不需要显示出来，如果用户选择“熟悉”那么只显示0和1的视频，如果用户选择“精通”那么0、1、2的视频都要显示。结构如表4.7所示：

表 4.7 video表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| vid | int(11) | NO | PRI | NULL | 视频的id |
| vname | int(11) | YES |  | NULL | 视频的名字 |
| cid | int(11) | YES | MUL | NULL | 视频所在课程id |
| tid | int(11) | YES | MUL | NULL | 任课老师id |
| url | Varchar(100) | YES |  | NULL | 视频所在路径 |
| tag | text | YES |  | NULL | 视频讲解知识点的关键词 |
| base | int(11) | NO |  | 1 | 根据用户选择学习这门课程的程度判断是否显示这个视频 |

1. user\_course表。

为了记录某个用户学习的课程信息的，便于系统后面使用协同过滤算法进行课程推荐，其中的user\_look\_num是记录用户学习这门课程看的视频数，这个是用户和所学课程的中间表。last\_look\_vid是记录用户最后一次观看这门课的视频的id，便于下次系统为用户定位，有更好的用户体验。其表结构如表4.8所示：

表 4.8 user\_course表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| uid | int(11) | NO | PRI | 0 | 用户的id |
| cid | int(11) | NO | PRI | 0 | 课程的id |
| score | int(11) | YES |  | 5 | 用户对课程的评分 |
| tid | int(11) | YES |  | NULL | 任课老师id |
| user\_look\_num | int(11) | YES |  | NULL | 用户看的视频数 |
| last\_look\_vid | int(11) | YES |  | NULL | 最后一次看的vid |

1. user\_question表。

为了便于使用Logistic Regression算法对用户进行评估，建立user\_question表记录用户做试题的正误情况，表结构如表4.9所示

表 4.9 user\_question表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| uid | int(11) | NO | PRI | 0 | 用户的id |
| qid | int(11) | NO | PRI | 0 | 问题的id |
| TOrF | float(2,1) | YES |  | NULL | 用户做题的正误，0是错误，1是正确 |

1. user\_subject表。

为了提高系统体验度，建立user\_subject表，记录每个user学习这个主题的进度，表中个字段记录如表4.10所示：

表 4.10 user\_subject表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| uid | int(11) | NO | PRI | 0 | 用户的id |
| sid | int(11) | NO | PRI | 0 | 主题的id |
| look | int(11) | YES |  | NULL | 用户是否学习过这个subject，0表示没有学习过，1表示学了但没学完，2表示已学习完 |

1. user\_teacher表。

记录了用户和老师相联系的信息。对于同一类课程，也许有不同的老师讲解，用户对不同的老师也有不同的评价，建立这个表示为了便于系统推荐用户更喜欢的老师的课程，结构如表4.11所示：

表 4.11 user\_teacher表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| uid | int(11) | NO | PRI | 0 | 用户的id |
| tid | int(11) | NO | PRI | 0 | 教师的id |
| score | int(11) | YES |  | 5 | 用户给老师的打分 |

1. user\_video表。

为了提高用户体验，建立user\_video表，记录用户是否查看了对应的video，便于系统展示给用户，结构如表4.12所示：

表 4.12 user\_video表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | Type | Null | Key | Default | 说明 |
| uid | int(11) | NO | PRI | 0 | 用户的id |
| vid | int(11) | NO | PRI | 0 | 视频的id |
| look | int(11) | YES |  | NULL | 用户是否学习了这个视频，0表示没有学习过，1表示学了但没有学完，2表示学习完了 |

# 系统实现与测试

## 各模块实现

通过对子模块的划分，本次研究使用MyEclipse2014作开发工具，对平台每个子模块都进行了实现，基本完成了平台所有功能。

### 登录注册模块

这个模块包含注册页面和登录页面。注册页面如图5.1所示。界面很简洁，包含一般注册该有的所有功能，包括用户输入格式合法性验证，另外，用户输入完之后，系统会对用户两次输入密码的一致性进行验证。如果注册成功，系统会给出跳转到如图5.2所示的登录功能的模块的选择。

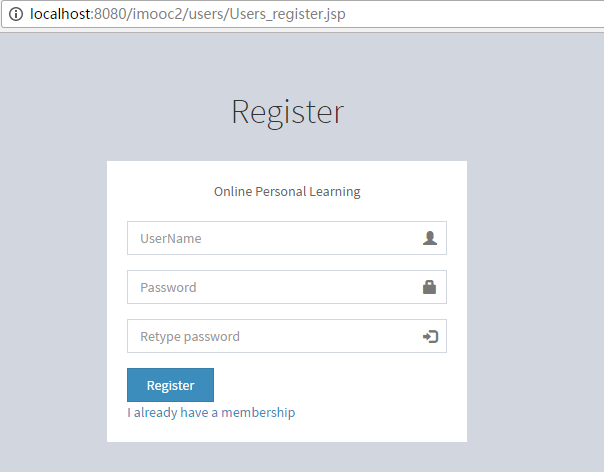


图 5.1 注册页面

登录页面如图5.2所示。包含基本的请求输入用户名、密码的输入框，同时下面通过使用cookie机制可以记住用户的输入，用户下次进入这个页面可以免于操作。另外，用户如果忘记用户名和密码，可以通过点击“I forget my password”进行重新设置，重设机制是：用户根据提示给出能识别自己的用户名，系统统一为用户分配“88888888”的初始登录系统的口令。

用户在进入到系统之后，系统会根据用户在数据有已有的数据进行一些初始化：

1. 如果用户之前已经加入了一些学习路径，系统会让用户跳转到“全部课程”页面，如图5.6，如果用户没有加入过学习路径，会跳转到“全部路径”页面，如图5.3。
2. 如果管理员新加入了一些课程，这些课程的关键词和某些主题关键词匹配，系统会根据AC自动机匹配课程的主题id

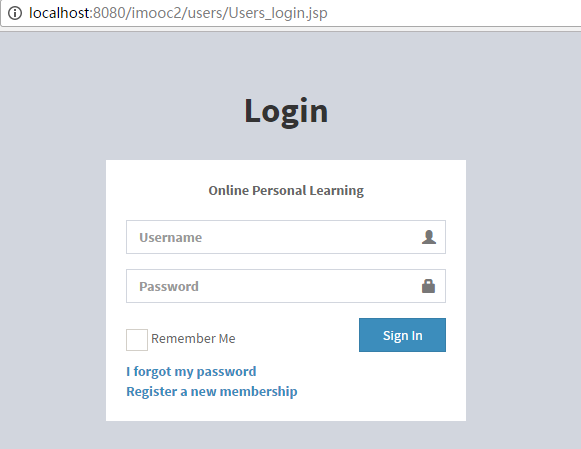


图 5.2 登录页面

### 课程资源模块

这个模块主要包括的功能有四个部分，分别是给用户提供学习路径接口的“路径选择”部分、让用户自主选择是否加入的“课程选择”部分、适应用户习惯生成的“课程推荐”部分和“课程评分”部分。

1. 路径选择。

路径选择页面如图5.3，这个页面会展示系统所有的学习路径。同时用户也可以点击“不用了，我自己随便看看。。。”跳转到“全部课程”页面进行自主选择课程。正如图示，如果用户之前学习过一个路径，系统会显示“继续学习”，如果没有学习过改路径系统会显示“加入学习”按钮。



图 5.3 全部路径页面

学习路径具体分阶段学习情况如图5.4，主要思路是每个学习路径分为多个阶段，每个阶段都有多个主题，每个主题包含多门课程，主题下面包含的课程如图5.5所示。



图 5.4 路径详细说明页面



图 5.5 主题包含课程页面

1. 全部课程。

“全部课程”页面显示的是学习系统包含的所有课程的链接，如图5.6所示，包含课程题目，课程基本介绍，课程的评分以及学习课程的人数等基本信息。

点击某一个课程进入，系统会显示课程所有章节，如图5.7所示，章节中用户对某些视频是否看过系统也会有标记，使用者在即将学习课程时，选择要对当前所要学习的课程吸收的程度。如果选择“了解”，只会显示数据库中course的base属性是1的章节视频，如果选择“熟悉”，会显示数据库中course的base属性是0和1的章节视频，如果选择的是“精通”，会显示该门课程所有的视频。

在课程最后，系统还会给出试题测试的入口，如图5.8所示：

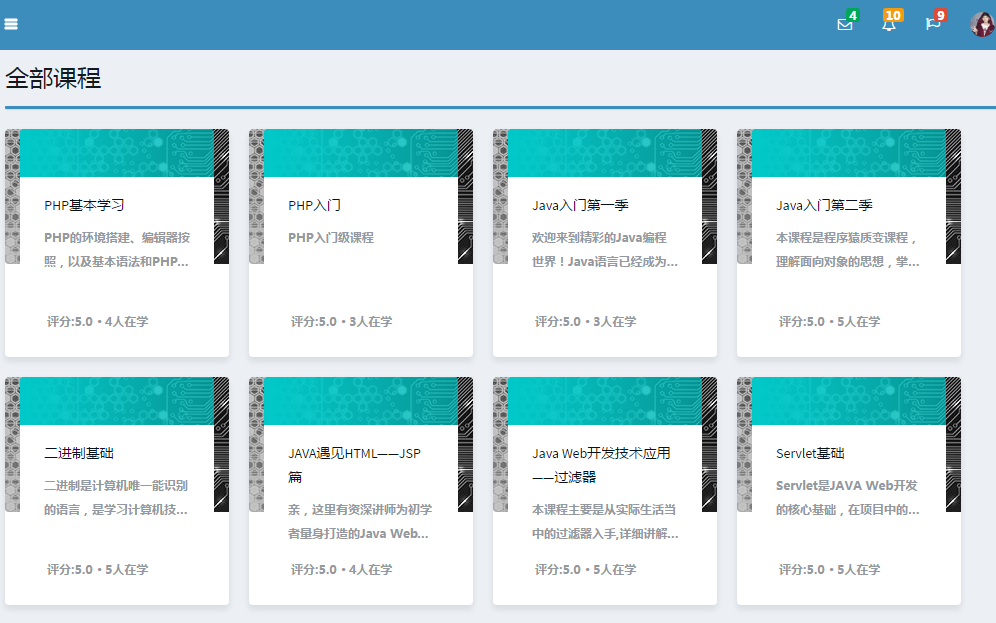


图 5.6 全部课程页面



图 5.7 course页面

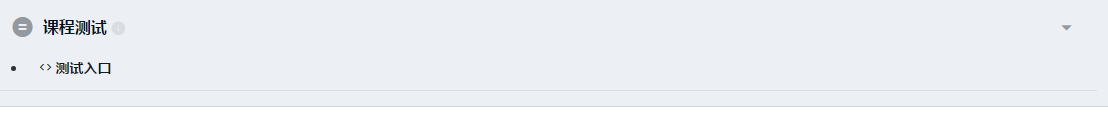


图 5.8 试题测试入口

1. 课程推荐

在这个页面，系统使用的是协同过滤算法为使用者显示没有学习过的课程，如图5.9所示：



图 5.9 课程推荐页面

这一方面的实现使用基于用户的协同过滤算法，借助Java的Mahout框架的Taste推荐引擎。从user\_course表中导出uid、cid和score字段的值存到.csv文件中，把.csv文件作为输入建立DataModel模型，然后调用相关接口算出当前登录用户uid和其他用户的欧几里得距离，根据系统设置好的邻居个数（假定为n）,取n个邻居作为推荐主题向当前用户推荐课程。

### 课程测试模块

1. 系统评估

这个模块是在用户学习完每门课程之后进入，目前系统实现的题目只有选择题，在进行做题之前，用户还要选择一次学习这门课程的目标(了解、熟悉或者精通)，这里不采用课程学习时用户选择的学习目标的原因在于，某些用户学习课程的目的和试题测试的目的可能不同。页面如图5.10所示。在提交了试题之后，根据用户提交的答案和标准答案的对比以及用户想要对这门课的掌握程度的选择，系统会给出测试是否合格的显示，如图5.11所示。

这个模块使用的是Logistic Regression算法，每次用户点击“提交”试题答案按钮之后，系统先判断用户每一题的正误，并把正误信息存到user\_question表，然后根据user\_question表的qid找出题目的diff字段信息，并把user\_question表中的TOrF数据筛选出来，把diff字段和TOrF字段的值存到一个txt文件中，作为算法的一个输入，并使用梯度下降法计算参数theta。

用到的公式如下：

|  |  |
| --- | --- |
|  | （5.1） |
|  | （5.2） |

这里的b为试题难度。

如果学生目标是“了解”那么参数b=3.5，如果是“熟悉”那么参数b=4.0，如果是“精通”那么参数b=4.8。求出的h为学生对于难度程度为b的题目求解正确的概率，如果概率大于0.7就说明合格，否则不合格。

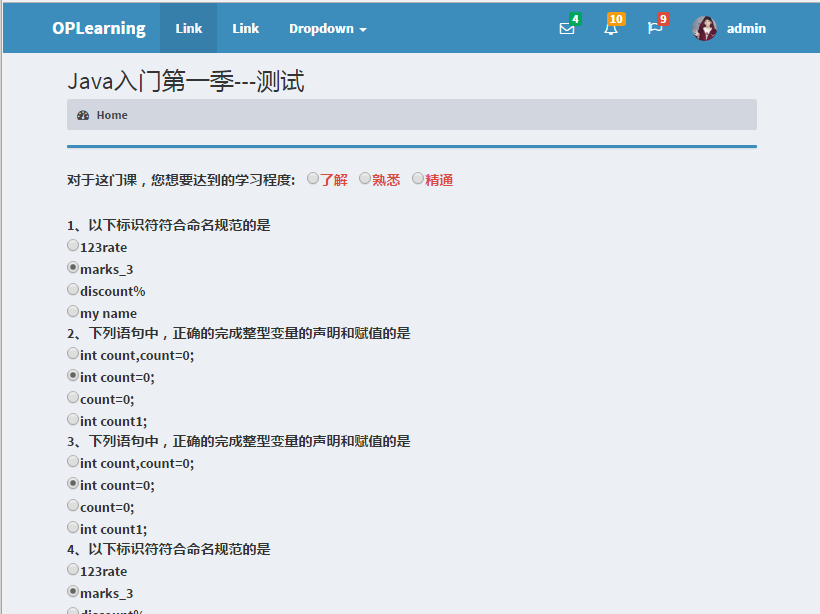


图 5.10 试题测试页面



图 5.11 试题评估页面

1. 快速复习

对于每一个试题，下面都会显示自己的keyword标签，利用AC自动机匹配算法，匹配对应知识点的视频，点击对应标签就自动跳转到那个视频，比如点击标签“标识符”他就会跳转到视频“认识Java标识符”页面，如图5.12所示。

## 系统测试

系统个性化主要体现在5个方面：提供标准学习路径、展示灵活的章节内容、进行更可靠的试题评估、试题快捷复习以及课程推荐。这里重点测试后四个方面。

1. 课程章节个性化

不同用户对于同一门课程想要达到的学习目的是不一样的，比如在“JAVA遇见HTML——JSP篇”这门课程里面，管理员设置了“静态网页与动态网页”和“搭建JAVA WEB开发环境”视频的base是0，设置了“Tomcat目录结构”和“WEB-INF目录详解”视频的base是2，如图5.13所示。当用户点击“了解”时，base为0和base为2的视频都被隐藏，如图5.14所示。当点击“熟悉”时，base为0的会显示，如图5.15所示。当点击“精通”时，第一章所有视频都被显示，如图5.16所示。



图 5.12 查看复习视频页面

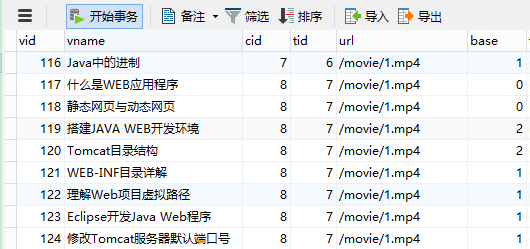


图 5.13 base为0和2的课程



图 5.14 点击“了解”章节显示



图 5.15 点击“熟悉”章节显示



图 5.16 点击“精通”章节显示

可以看到图5.14中，base是0或者2的视频（第1到4节内容）都被隐藏，

图5.15中，base是2的视频（第3和第4节内容）被隐藏，图5.16中，第一章内容全部显示。

1. 试题评估个性化

用于测试的题目有39个，用户admin的uid是3，在课程“Java入门第一季”中，他做题的正误情况如表5.1所示：

表 5.1 uid=3做题正误表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 难度 | 正误 | 难度 | 正误 | 难度 | 正误 |
| 2.9 | 1.0 | 2.0 | 1.0 | 5.0 | 0.0 |
| 1.9 | 1.0 | 3.5 | 1.0 | 2.2 | 1.0 |
| 0.9 | 1.0 | 4.0 | 0.0 | 3.2 | 1.0 |
| 3.0 | 1.0 | 2.9 | 1.0 | 4.2 | 1.0 |
| 4.0 | 1.0 | 1.9 | 1.0 | 3.4 | 1.0 |
| 2.5 | 1.0 | 0.9 | 1.0 | 1.0 | 0.0 |
| 1.5 | 1.0 | 3.5 | 1.0 | 1.0 | 1.0 |
| 0.5 | 1.0 | 4.2 | 0.0 | 2.0 | 1.0 |
| 3.0 | 1.0 | 3.5 | 1.0 | 4.0 | 0.0 |
| 4.0 | 1.0 | 2.2 | 1.0 | 1.0 | 1.0 |
| 2.0 | 1.0 | 3.5 | 1.0 | 2.0 | 1.0 |
| 3.3 | 1.0 | 4.2 | 0.0 | 3.0 | 1.0 |
| 3.7 | 1.0 | 3.2 | 1.0 | 4.0 | 0.0 |

使用算法计算h在控制台输出为：

h= 0.7267796590291893

页面显示如图5.17所示：

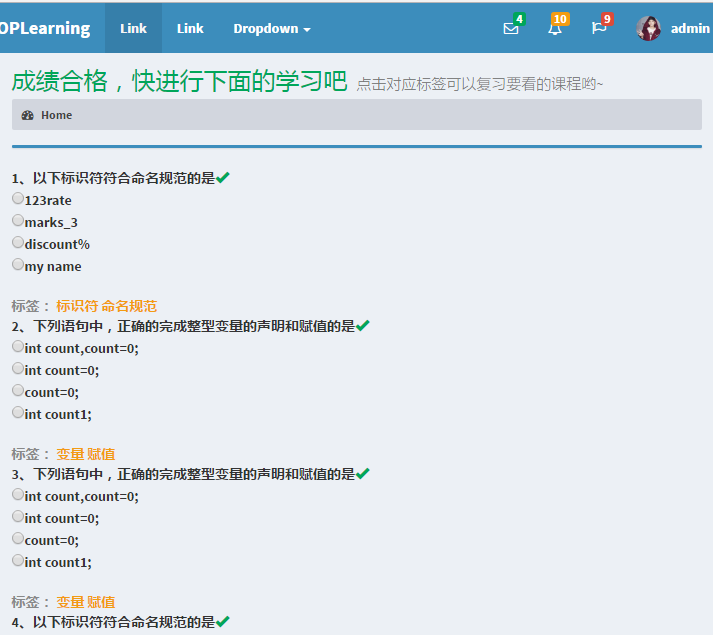


图 5.17 用户提交后的评估结果

对于同样一组数据，如果用户选择“熟悉”，控制台测试输出的h是：

h= 0.5602788055583934

页面显示如图5.18所示：

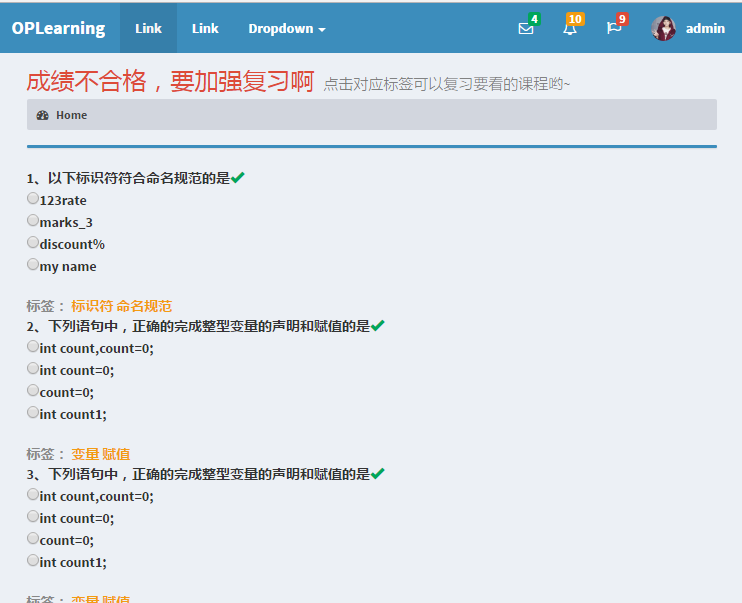


图 5.18 用户提交后的评估结果

同理，用户选择“精通”，测试得到的h是：

h= 0.28183285703624056

页面输出同图5.18。

由此可见，系统的Logistic Regression算法不是单单地通过计算用户做题正确的多少来判断用户的能力，而是综合考虑题目难度、用户想要达到的学习课程的目的等因素综合评估用户是否达标，这样的评估结果更有说服力。

1. 试题快捷复习个性化

试题快捷复习是使用AC自动机匹配算法便于用户快速复习。如对于“Java入门第一季”这门课，对应章节视频的id为62~67的tag关键词如图5.19所示，试题在数据库中的存储如图5.20所示。

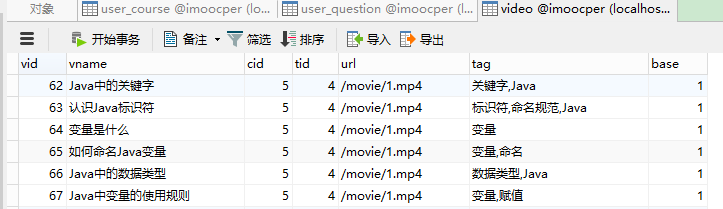


图 5.19 部分视频表数据



图 5.20 部分question表数据

评估过后，可以看到每个问题下面都有自己的标签，用户点击第一题的标签“标识符”，预期应该是匹配到vid=63的视频，页面显示如图5.21所示：

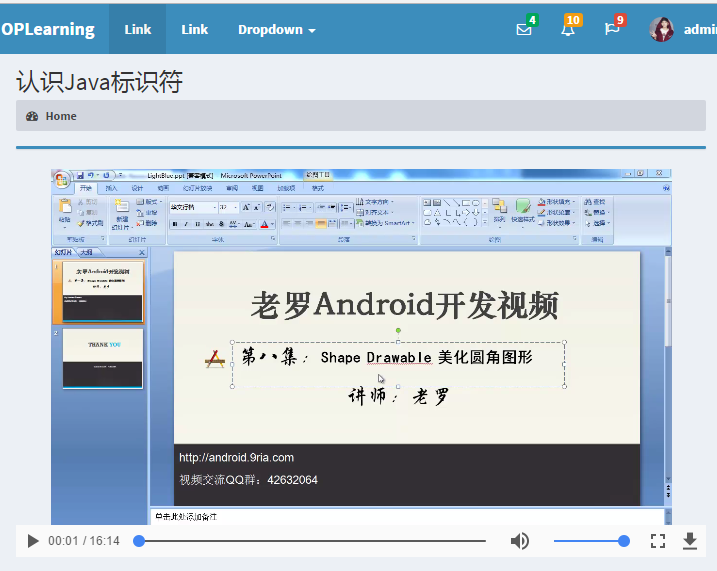


图 5.21 复习页面

发现确实是vid=63的“认识Java标识符”页面。

同理当用户点击第2题的“变量”标签，预期显示vid=64的视频，页面显示如图5.22所示：

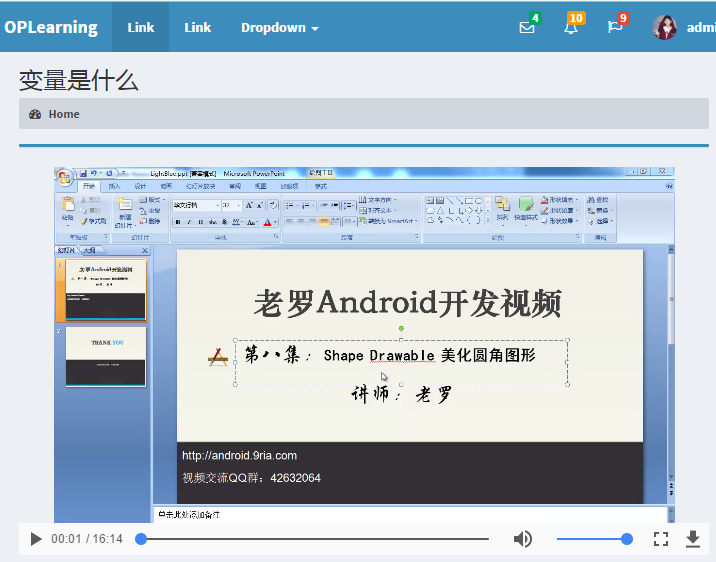


图 5.22 复习页面

通过上面的测试，系统满足用户通过试题的keyword快速定位要复习的课程的目的，这种实现方式对于提高用户体验有很大帮助。

1. 个性化推荐

测试数据如表5.2所示：

表 5.2 user\_course表部分记录

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| uid | cid | score | uid | cid | score |
| 3 | 2 | 5 | 5 | 4 | 4 |
| 3 | 7 | 3 | 5 | 6 | 4 |
| 3 | 9 | 5 | 6 | 5 | 0 |
| 3 | 10 | 4 | 6 | 2 | 3 |
| 3 | 12 | 4 | 6 | 9 | 3 |
| 3 | 5 | 5 | 6 | 6 | 4 |
| 3 | 8 | 3 | 6 | 14 | 3 |
| 4 | 6 | 4 | 6 | 7 | 5 |
| 4 | 12 | 2 | 6 | 10 | 1 |
| 4 | 4 | 4 | 7 | 9 | 3 |
| 4 | 10 | 2 | 7 | 7 | 5 |
| 4 | 14 | 5 | 7 | 10 | 3 |
| 4 | 8 | 3 | 7 | 6 | 4 |
| 4 | 2 | 3 | 8 | 10 | 2 |
| 5 | 7 | 4 | 8 | 14 | 4 |
| 5 | 8 | 2 | 8 | 9 | 2 |
| 5 | 14 | 3 | 9 | 13 | 3 |
| 5 | 9 | 4 | 9 | 4 | 5 |
| 5 | 13 | 3 | 9 | 7 | 4 |
| 5 | 5 | 5 | 9 | 6 | 3 |
| 5 | 2 | 2 | 9 | 14 | 5 |

把该记录作为系统推荐算法的输入，系统设置最多推荐项目是3，控制台输出如图5.23所示：

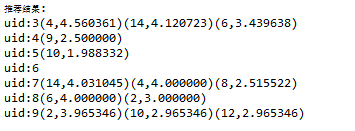


图 5.23 系统所有用户推荐

可以看到系统为每个用户都推荐了对应项目的cid。系统也有预期显示。

# 总结与展望

## 工作总结

Mooc平台由于其自身所具有的特性，包括无时空限制、注册门槛低和态性和开放性强等，吸引了众多用户。随着Mooc平台资源的不断增多，如何为用户选择个性化的学习内容、建立个性化的学习过程逐步成为了一个突出的问题。本文针对Mooc平台在个性化内容定制方面存在的典型问题，提供了一个综合的解决方案，主要的设计要点包括：

1. 系统除了提供供用户自由学习的视频课程之外，还为很多领域提供了丰富的学习路径，便于用户更加系统地学习某个领域。
2. 采用基于用户的协同过滤算法为用户推荐可能想要学习的课程，有效地挖掘用户潜在的学习兴趣。
3. 对于同一门课程，不再是固定地给出不变的学习章节，而是根据用户选择对这门课程想要的了解程度灵活的给出学习内容。
4. 在学习完一门课程之后，给一定量的测试题，评估方法采用Logistic Regression模型对用户的答案进行评估，结合用户答案的对错情况和用户想要对这门课的了解情况给出更加可靠的及格与否的评价。
5. 在课程测试方面，除了第（4）点的评估方面，还在每个题下面给出每个题目的知识点关键词链接，链接到对应课程的讲解，便于用户快速复习。

## 研究展望

由于时间和个人知识程度的限制，本系统在个性化方面的实现还是比较浅层的，在个性化学习的实现方面本次研究还有很大的提升空间：一、系统学习资源方面，由于本次研究只是一个测试系统，学习资源还不够丰富，不具有多样性；二、本系统只实现了“视频”一种学习方式，对于某些领域比如计算机编程方面增加在线编程模块是很有必要的；三、在试题测试方面，本系统只实现了选择题这一种题型的测试。

针对以上几点，对系统进行优化迭代将成为接下来的工作重心。

参考文献

1. 杨勇. 浅谈计算机网络管理技术的发展趋势[J]. 消费电子, 2014(22):102-102.
2. 韩洪举. 杜威现代教育思想解析[J]. 中国成人教育, 2001(2):50-51.
3. 蒋东兴．MOOC摸到了“象鼻子”U教育网络， 2015(4)：25．
4. 吴明华.“大规模开放在线课程”的兴起[J]. 世界知识，2013（14）60-61.
5. 李武静. 美国社区学院开放教育资源运动研究[D]. 厦门大学, 2012.
6. 孙志雷. 个性化网络学习环境的设计与开发——以《计算机辅助教学》为例 [D]. 南京师范大学, 2006.
7. 史龙珍，韩小飞. 基于 MOOC 的个性化学习模式研究[J].软件导刊 ,2014(6): 185-187
8. 王怡涵. 能力导向终身学习个性化推荐系统学习者模型构建[D]. 东北师范大学, 2011.
9. 冯英伟. 基于智能Agent技术的网上教学系统的研究与设计[D]. 沈阳师范大学, 2007.
10. 黄伯平. 基于语义网技术的个性化学习服务机制及功能建模研究[D]. 东北 师范大学, 2010.
11. Oldale A, Oldale J, Reenen J V, et al. COLLABORATIVE FILTERING: US, US 20040054572 A1[P]. 2004.
12. 程琳. 主流MOOC平台对比分析研究[D]. 南京邮电大学, 2015.
13. 何彪. 面向MOOC的在线学习平台的设计及应用研究[D]. 大连理工大学, 2015.
14. 吴筱萌, 雍文静, 代良,等. 基于Coursera课程模式的在线课程学生体验研究[J]. 中国电化教育, 2014(6):11-17.
15. 李海琴. 美国edx平台调查报告分析慕课8大特点[J]. 中国教育信息化, 2014(14):77-78.
16. 约翰·丹尼尔 ，王志军。让 MOOCs 更有意义：在谎言、悖论和可能性的迷 宫中沉思[J]. 现代远程教育研究，2013（3）：2-12,27.
17. 苑宁. 知乎·果壳:知识型问答社区对比研究[J]. 新闻研究导刊, 2015(17):258-259.
18. 李静, 王亚琴, 张仁东. 网络环境下医学院教与学的转变[J]. 中外健康文摘, 2013(33):239-240.
19. 杨文建, 邓李君. 我国主要MOOC平台现状分析及发展策略[J]. 重庆第二师范学院学报, 2016, 29(5):168-173.
20. 朱宝玉. 网络课程中自适应学习技术的研究[D]. 东北大学, 2006.
21. 李华, 何茜, 吴中福. 基于Web的个性化学习系统研究[J]. 计算机工程与应用, 2002, 38(13):239-242.
22. 黄悦. 终身学习与个性化学习[J]．科教导刊:中旬刊，2013（7）：11-12.
23. 樊春运. 个性化网络学习环境设计原理及实现[D]. 东北师范大学, 2006.
24. 邵海珠. 基于协同过滤的疾病基因预测方法[D]. 西安电子科技大学, 2014.
25. Konstan，J．，Miller B．，Maltz D．，Herlocker J．，Gordon L．，Riedl J．Grouplens： Applying Collaborative filtering to Usene!t News，Communications of the ACM，V01．40(3)，pp77-87，1996．
26. 陈衍仪，图像的分形理论和方法，1997，国防丁业出版社，北京，14—36，86—99
27. 刘静. 多维度可扩展推荐引擎框架的设计与应用[D]. 哈尔滨工业大学, 2013.
28. M. R. McLaughlin and J．L．Herlocker,A collaborative filtering algorithm and evaluation metric that accurately model the user experience，in Proceedings of 27th Annual International ACM SIGⅡL Conference Oil Research and Development in Information Retrieval(StOre’04)，PP．329\_336，Sheffield,UK,2004
29. 夏培勇. 个性化推荐技术中的协同过滤算法研究[D]. 中国海洋大学, 2011.
30. 龚松杰. 个性化推荐中一种新的相似性计算方法[J]. 计算机系统应用, 2008, 17(7):89-91.
31. 皮佳明. 基于用户兴趣变化的协同过滤推荐算法研究[D]. 云南财经大学, 2014.
32. 李雷. 基于社会化标签的协同过滤个性化推荐策略的研究与实现[D]. 东北大学, 2014.
33. 朱倩. 浅谈基于Mahout推荐引擎的构建[J]. 数字技术与应用, 2015(4):44-45.
34. 李龙飞. 基于Hadoop+Mahout的智能终端云应用推荐引擎的研究与实现[D]. 电子科技大学, 2013.
35. 于嘉. 基于MAHOUT的几种推荐算法的组合实现与评测[D]. 华中师范大学, 2015.
36. 内特. 应用线性回归模型[M]. 中国统计出版社, 1990.
37. 张利萍, 李宏光. 灰色神经网络预测算法在DMF回收过程中的应用[J]. 微计算机信息, 2005(1):183-184.
38. 张萧, 黄晞, 仲伟汉,等. Sigmoid函数及其导函数的FPGA实现[J]. 福建师大学报(自然科学版), 2011, 27(2):62-65.
39. 刘颖超, 张纪元. 梯度下降法[J]. 南京理工大学学报自然科学版, 1993(2):12-16.
40. D Gusfield Algorithms on Strings,Trees,and Sequences.Cambridge University Press. 1997.
41. NAVARRO G，RAFFINOT M.Flexible Pattern Matching in Strings：Practical On.1ine Search Algorithms for Texts and Biological Sequences[M].SIAM Publishing，2002．
42. 张坤. 基于深度包检测的网络应用识别与流量控制[D]. 广东工业大学, 2016.

# 致谢

本次课题的研究历经几个月，是自己努力专研的结晶，更多的是来自老师、朋友和家人的支持。

首先，衷心地感谢尊敬的导师何扬帆教授，本次研究工作是在她的指导下完成的，这里的每一份成果都得益于老师的关心和建议。何老师是位认真负责的导师，整个课题的研究老师都一丝不苟地对我进行监督和指导，感谢老师的付出！

接下来要感谢的是我的父母，感谢一路走来他们对我的陪伴和支持，作为子女时刻能感受到他们对我的疼爱，我的每一个细微的进步都离不开他们的支持。

最后要感谢的是我大学四年的老师和同学，在各位老师的关心和培育下，以及各位同窗的帮助下我度过了充实忙碌的大学生活。在学校四年经历从生活到学习上的进步，这里每一步的跨进都要感谢老师对我的教育和关爱，每一个经历也离不开同学们一路的陪伴和鼓励。