
基于项目反应理论的计算机辅助 测验系统的研究与应用

摘 要

在教学过程中，为全面了解学生对于课程知识点的掌握情况，通常采用试题的形式进行考核，而传统纸笔形式的考核模式缺乏灵活性，对于不同能力的考生提供相同的测试项目，无法根据学生在测试过程中反映出来的能力水平灵活地为学生提供与其水平相适应的测试项目，在测验过程中存在着学生遇到较多与自身能力不匹配的测试项目的情况，致使最终测验结果与自身真实能力值存在较大偏差的情况。这种情况不利于教师对学生知识点掌握程度进行评估，从而调整其教学计划提高教学质量。本文针对现有测验模式的不足，引入项目反映理论的内容，并设计了基于 Logistic 单参数模型的计算机辅助测验系统，建立强大的试题库，使该系统可以根据不同学生的能力水平，向其提供能力匹配的测试项目，使测验结果更能反映学生的真实水平。

关键词：项目反映理论；计算机辅助测验系统；Logistic

Abstract

In the teaching process, in order to fully understand the students' mastery of the course knowledge points, the test questions are usually used in the form of examination, while the traditional paper-pen test forms are lack of flexibility, providing the same test items for candidates of different abilities, not able to flexibly provide test items that are compatible with their own abilities performed in the testing process. Students encounter more test items that do not match their own abilities, leading to relatively large deviation between the test results and their own true ability value. This situation is not conducive to teachers to assess the degree of mastery of students' knowledge points, thus adjusting their teaching plans to improve the quality of teaching. In view of the shortcomings of the existing test mode, this paper introduces the content of the item response theory (IRT), and develops a computer-aided test system based on the logistic model and builds a powerful test question database, so that the system can provide the students with compatible test items according to the ability level of different students, therefore the test results can better reflect the true level of the students.

Keywords: Item Response Theory; computer-aided test system; logistic mode

目录

Abstract.....	II
1. 引言	1
1.1 题目背景及意义.....	1
1.2 本研究国内外应用研究.....	1
1.3 论文内容与组织.....	2
2. 教育测量的理论研究回顾	3
2.1 经典测量理论 (CTT)	3
2.1.1 经典测验理论的局限性.....	3
2.2 项目反应理论 (IRT)	4
2.2.1 项目反应理论模型.....	5
2.2.2 项目反应理论的优点	6
2.2.3 基于项目反应理论的计算机辅助测试系统.....	6
3. 基于项目反应理论的计算机辅助测验系统设计	8
3.1 系统设计架构图.....	8
3.2 数据库设计.....	9
3.2.1 数据库集合设计.....	9
3.3 题库的组成.....	12
3.3.1 试题库的构建.....	12
3.3.2 题目类型的选择.....	12
3.3.3 题目参数的确定.....	12
3.4 系统关键模块的实现	13
3.4.1 被测个体初始能力值估计模块.....	13
3.4.2 自适应选题模块.....	15
3.4.3 能力参数模块.....	16
3.4 运行环境.....	18
3.5 开发技术.....	19
3.5.1 Vue 框架简介.....	19
3.5.2 Element 框架简介.....	19
3.5.3 Express 框架简介.....	20
3.6 系统界面	20
致谢	25
参考文献:	26

1. 引言

1.1 题目背景及意义

人类通过测量的方式来认知这个世界,通过测量活动,人类可以知道事物的发展规律甚至还能根据已有的测量结果进行相应的决策或者改变某种行为,致使最终效益最大化或者损失最小化,达到理想的效果。这种行为存在于各种领域,教学领域中的核心环节之一就是教育测量,通过教育测量对教育及其造成的行为变化进行数量化的确定。教育的目的在于通过教育的方式改变接受教育的个体的行为方式。由于个体心智上的差异,接受相同的教育,对不同个体的行为上的影响也是不同的。教育测量的核心内容就是对学生个体的学业水平、思想素质等方面等给出科学的、定量的表述,而目前教学测量的方式仍比较单一,在教学过程中教师通常使用传统的纸笔测验的形式来考查学生的学业掌握情况,之所以采用这种形式并不是这种方式是最好的,这种方式也存在着弊端,即“千人一卷”——所有被测个体都采用先用的测试项目进行测试,无法根据针对不同的个体在测试过程中表现出来的能力水平动态地出卷,导致测量结果与被测个体的真实水平存在偏差。

教育测量得到的结果,可用来评价、指导教育行为,选拔人才等^[1]。测量结果的重要性不言而喻,它侧面上对测量方式提出了更高的要求以求更加准确的测试结果。计算化自适应测验的出现能有效解决“千人一卷”的问题,做到“因人施测”,更科学准确的测量被测个体的真实水平。

1.2 本研究国内外应用研究

随着计算机技术的飞速发展以及项目反应理论的日趋完善,计算机自适应测验从理论转变成现实。

国外方面,20世纪80年代起,计算机自适应语言测试在美国,主要应用于分级测试以及能力测试两个领域,例如美国杨百翰大学开发的第一个适应性语言测试就是用于分级目的,它包括了法语、德语、西班牙语和俄语等不同语种的测试^[2]。在此适应性语言测试诞生之后,大规模考试如GRE,TOFEL等先后采用计算机适应性测试。除了之外,美国教育考试服务中心(ETS)还将计算机自适应测试应用于研究生资格考试、护士资格证书考试以及建筑师资格考试等。

国内方面,江西师范大学研制的基于项目反应理论(IRT)的小学阶段数学课程自适应测试系统以及中学英语水平计算机自适应测试系统都取得了较为理想的应用效果;2002年3月份初,由南京师范大学主导研发的化学自适应测验系统在网络上投入运行;2016年由山东师范大学主导研制的大学英语六级词汇自适应测试系统投入使用,从山东省济南市某高校中随机选取非英语专业的112名大二年级的学生作为被试,反映良好。这些高校在自

适应测试理论方面研究取得了进展并研发了相应的自适应测试系统，但这些系统的使用规模较小，受众少，未能够应用于大规模的能力水平测试中。国内计算机自适应测验系统还需要进一步的研发和推广。

1.3 论文内容与组织

本文主要研究的是项目反应理论在计算机辅助测试上的应用，研发以项目反应理论为依据的辅助测试系统，实现“因人施测”，主要包含智能组卷、试题发布、个人记录等功能。本文同时对于开发过程中遇到的一些问题进行分析、总结。本文按照组织结构划分本文可一共分为六章：

第一章，引言。主要介绍本文进行该题目研究的背景及意义，同时介绍国内外对于项目反应理论在自适应测试中的应用现状，并以此提出基于项目反应理论的计算机辅助测验系统的研究与应用。

第二章，教育测量理论研究回顾。大体介绍了教育测量理论的重要理论内容——经典测验理论以及项目反应理论，并对这两个理论加以分析。

第三章，主要是对基于项目反应理论的计算机辅助测验系统进行设计与研究，对框架设计流程进行解释，同时对使用的技术以及算法进行分析说明。

第四章，主要是对设计开发过程中遇到问题进行思考和总结，以及使用到一些解决方法，还有一些开发技巧。

第五章，主要是对本计算机辅助测验系统的工作进行总结以及对项目反应理论在计算机自适应测试系统上的展望。

2. 教育测量的理论研究回顾

测量的实践离不开理论的指导。教育测量理论经过多位大师几十年的研究之后愈发变得成熟，其中经典测量理论（Classical Test Theory, CTT）和项目反应理论（Item Response Theory, IRT）属于就教育测量理论中的两大类。教育测量理论的发展可分为两个阶段，第一个阶段是 19 世纪 50 年代以前的经典测量理论一枝独秀阶段，当时多数适应性测试基于经典测量理论来进行；第二个阶段便是以项目反应理论占主要地位的多种测量理论并存时期，该阶段中项目反应理论逐渐成熟，由于项目反应理论的强大在加上自身的不足，经典测量理论逐渐显得暗淡，被项目反应理论所取代，后者成为自适应测试中的研究重点。

2.1 经典测量理论（CTT）

经典测量理论是最早发展起来的教育测量理论，亦叫真分数理论。经典测验理论基于行为主义的刺激反应理论，在测验中，要求被试完成一定的作业或试题，然后根据其实际作答的表现来推论和评价其心理特质的发展水平^[3]。基于该理论研发的测试系统其测量分数 X 实际上由两部分组成，其一是被测个体的所测特性的真实值 T （即真分数），另一部分则是误差分数 E ，数学表达式为 $X = T + E$ 。经典测验理论是基于弱假设的，能够成立的前提是必须满足以下 3 个假设：

- (1) 假设真分数表示的是被测个体的特质，在一定程度上保持恒定不变，且为一个常数。
- (2) 假设误差是完全随机的。一方面是要期望值是是 0 的正态随机变量，另一方面要求测量误差是完全独立的随机变量，与真分数互不影响，相互独立
- (3) 假设真分数与误差分数相加之和即为观察分数的值。

2.1.1 经典测验理论的局限性

在现代测量理论出现、成熟之前，经典测量理论教育测验的研制起到过相当大的指示作用，但是它仍存在许多的不足之处。经典测验理论的局限性主要体现在：

(1) 样本的依赖性：被测群体的会对经典测验理论中测试项目的难度、区分度等指标造成影响。在被测群体普遍水平较高的情况下，会测得难度值较小，反之则会造成难度值较大。假如被测群体能力较为接近，将导致个体之间分差不大，测验的区分度较小，反之则会导致区分度较大。

(2) 能力量表与难度量表不一致：被试能力量表是测验的卷面总分，项目的难度量表是被试群体的得分率。被试个体卷面总分的参照系是本次测验中的所有项目，被试得分 60 表示该被试正确作答了全部项目的 60%，然而项目难度量表的参照系是被试群体，项目难度

0.60 表示有 60% 的被试个体正确作答了该项目。由于两个量表的参照系完全不同，因此我们无法找到与被测能力水平恰好匹配的测试项目。

(3) 信度理论的不完善：CTT 使用信度来表示测验的准确性和可靠性，其理论定义为：一组考试分数中真分数方差与观察分数方差的比率。然而在实际操作中难以实现严格平行的测验，因此只能假设测验是严格平行的，才能对信度进行取值。

(4) 无法预测正确反应概率：CTT 中能力量表与难度量表使用的参照系不同，不仅导致无法动态地提供与被测个体能力匹配的测试项目，还导致无法对被试个体作答新的测验项目上正确的概率进行预测。

中国传统纸笔形式的测验方式就是以经典测验理论为基础的，对于所有的被测个体提供一模一样的测试内容并按要求被测个体在规定时间内完成，无法做到“因人施测”。正如前面提到的 CTT 的不足，这种形式的题目的指标很大程度上依赖于样本。简单的题目大部分被测个体都能答对，较难的题目大部分都难以答对，而且无法把测试难度与被测个体的能力水平统一到同一量表上，动态地提供与被测个体能力匹配的测试项目，导致难以测出被测个体的真实水平。

2.2 项目反应理论（IRT）

项目反应理论（Item Response Theory, IRT）是现代测量理论中影响力最大的测量理论，与概化理论一起改变了经典测验理论一统测量领域的格局，也称潜在特质理论或潜在特质，是心理统计学模型的总称，针对经典测量理论的缺陷，能够克服经典测量理论的局限性^[4]。IRT 诞生之初由于理论比较复杂加上操作难度大等缘故限制了它的使用，但由于计算机技术的飞速发展，在测验中运用 IRT 变得不再那么困难。

项目反应理论的核心是确定被试能力水平与被试对项目的应答结果之间的关系，用数学思维确定的这种模型便是“项目反应模型”，所以准确地建立这个数学模型，并对各个参数进行准确而有效的估计是项目反应理论的重要部分^[4]。

区别于 CTT，IRT 是建立在强假设的基础之上的。项目反应理论是建立在“能力单维性假设”“局部独立性假设”和“项目特征曲线假设”三个基本假设基础上^[5]。

(1) 能力单维性假设：所有测验项目均是为了测量被测个体某个潜在的特质，然而在测验过程中测验本身无法排除其他特质对测试结果的影响，因此只要满足该潜在特质在测验过程中对被测个体的行为选择占据主要作用即可。

(2) 局部独立性假设：测验过程中被测个体的反应不受其他测试项目的影响，只与当前测试项目有关。

(3) 项目特性曲线假设：由于项目参数和能力参数具有不变性，这种参数不变性决定了每一个项目都有各自唯一的一条非线性项目特征曲线，特定项目的特征曲线的形状和位置都是固定的，从曲线上所映射出的各个变量之间相互独立，即曲线上所反应出的项目参数和被试能力参数互不影响^[4]。

2.2.1 项目反应理论模型

项目理论模型是 IRT 的重要部分，目前已知的模型种类不下 20 种，其中最常用的是 Logistic 模型。按照参数的个数，Logistic 模型及相应的项目特征函数可分为以下三种：

$$\text{单参数模型: } P(\theta) = \frac{1}{1+e^{-D(\theta-b)}} \quad (2-1)$$

$$\text{双参数模型: } P(\theta) = \frac{1}{1+e^{-Da(\theta-b)}} \quad (2-2)$$

$$\text{三参数模型: } P(\theta) = c + (1-c) \frac{1}{1+e^{-D(\theta-b)}} \quad (2-3)$$

其中 a 、 b 、 c 分别表示测试项目的区分度、难度、猜测系数； D 为一个常数，取值为 1.7； $P(\theta)$ 表示能力为 θ 的被测个体答对该测试项目的概率。

在 IRT 的众多模型里,最常用的是 Logistic 三参数模型，其项目特征曲线（Item Characteristic Curve, ICC）的示意图如下：

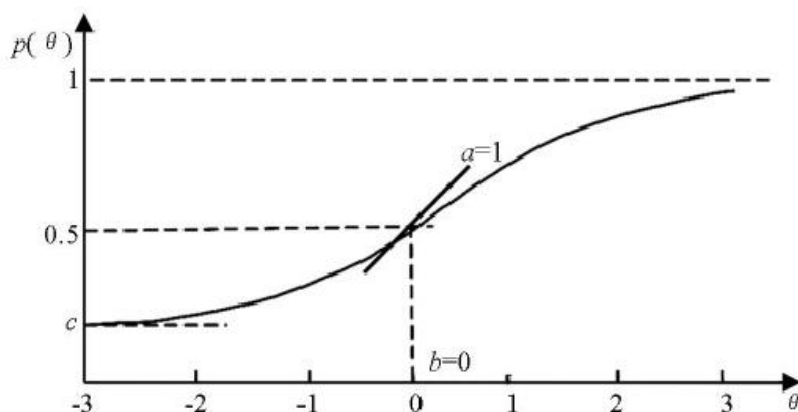


图 2-1

由图 2-1 所示，我们可知：

- (1) ICC 在纵坐标轴上的截距（即猜测参数 c ）表示在被测个体能力水平取值范围为 $(-3,3)$ 的情况下，即使在被测个体能力水平极低——能力值接近 -3，也有可能通过猜测的方式对该测试项目做出正确作答。
- (2) ICC 在横坐标轴上的投影为题目难度 b ，当 $P(\theta) = 0.5$ 也即曲线斜率达最大时 $b = 0$ ，此时，表示项目的难度适应大多数被试的能力^[6]；
- (3) ICC 的斜率表示的是项目的区分度 a ，斜率越大， $P(\theta)$ 对能力水平的变化就越敏感，则该测试项目对被测群体的区分度越大。

三参数模型综合考虑了测试项目的区分度、难度以及猜测系数，因此能够提供的信息量更大，测验结果也更加准确，通过相对来说三参数模型参数的处理、计算也更加繁琐。考虑到单参数模型较为简单，易于实现，为降低系统的开发难度，本系统设计使用的是 Logistic 单参数模型，只考虑项目难度 b 对被测个体能力 θ 的影响。

2.2.2 项目反应理论的优点

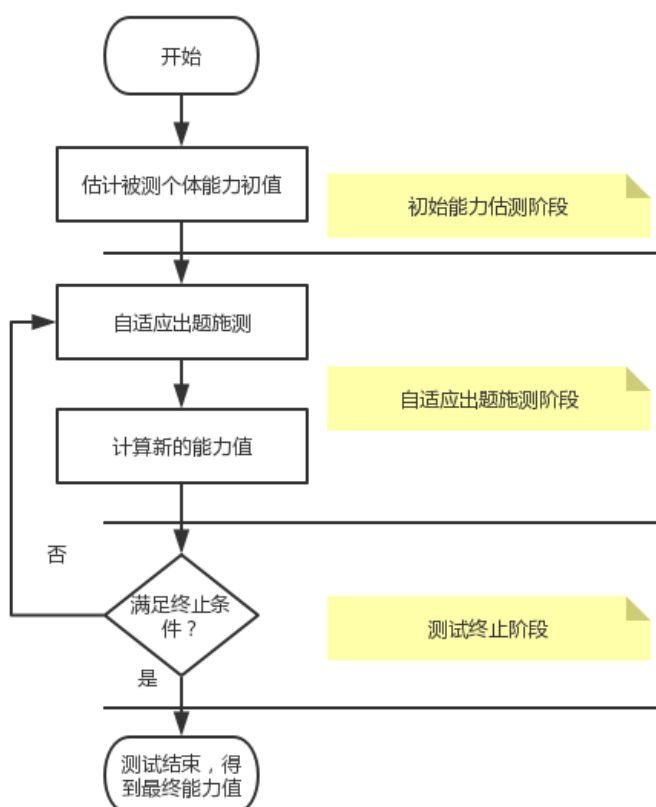
作为现代测量理论之一，项目反应理论能够很好地克服经典测量理论的局限性，其优点包含以下 3 点：

- (1) 参数不变性：区别于 CTT，IRT 参数的估计独立于被测样本，取值于被测样本无关。参数一旦确定就不再变化，这是 IRT 最大的优点。
- (2) 量表一致性：被测个体的能力和测试项目的参数都是定义在同一个维度的量表上
- (3) 可预测性：由于量表一致性，因此使用相应的算法就可以算法被测个体在新的测试项目上作答正确的概率。

目前 IRT 已被广泛应用于心理学、教育诸多领域的大规模的自适应测验中，例如 GRE、TOFEL 等。现代测量理论的兴起，导致了经典测验理论渐渐退出测量领域的历史舞台。

2.2.3 基于项目反应理论的计算机辅助测试系统

借鉴 CAT 通用流程，在考试初始阶段先获得被测个体的初始能力值，这个能力值是粗略的，接下来根据被测个体的答题情况不断的进行估计、修正，并动态的根据被测个体的能力值从题目选择合适的题目，直至达到终止条件时即可结束考试，得到被测个体的能力值。整个测验过程大致可分为 3 个阶段：初始能力估测阶段、自适应出题施测阶段以及测试终止阶段。



（1）初始能力估测阶段

根据 CAT 过程，对被测个体能力进行估值是重要的一个环节，原则是估值越准确越好。能力估测方法可以结合实际情况进行选择，可能的估测方式有以下 3 种选择。第一种选择是被测个体在该系统不存在任何的答题记录的情况下，系统先提供几道难度中等的题目，根据被测个体的答题结果进行计算得出被测个体的能力值；第 2 种是被测个体在系统中有答题记录，那么直接使用数据库中相应的能力值作为初始能力值即可；第 3 种方式是出题者根据对被测个体的了解，直接估值。能力估值结果需要存入数据库供后续阶段使用。

（2）自适应出题阶段

读取系统中被测个体的能力值，使用选题策略在题库中进行选题施测，并根据被测个体的作答结果更新系统中被测个体的能力值，在不满足测验终止条件时循环此步骤。

（3）测试终止阶段

满足终止条件，测试结束，保存结果到数据库。

3. 基于项目反应理论的计算机辅助测验系统设计

3.1 系统设计架构图

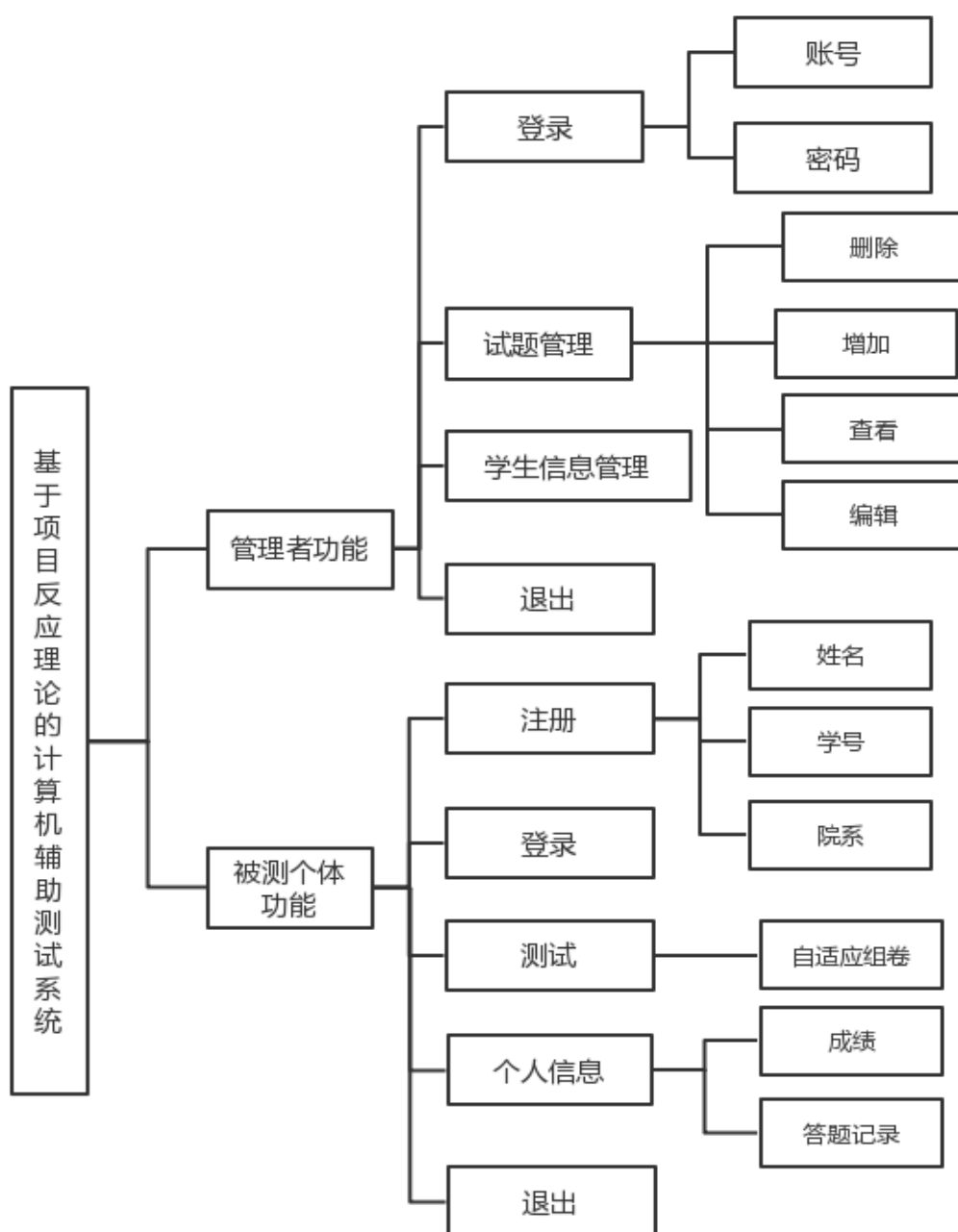


图 3-1 系统架构设计图

在本系统设计中，用户可以通过前端页面进行登录，系统会自动辨别用户身份是管理员还是被测个体并展示不同的页面内容，能够进行的操作也不同，管理员能够执行更

多操作。若为管理员，则可以进行试题的管理，如增加试题、编辑试题等；若为被测个体，则可以查看自己的成绩、进行考试、答题记录等。更多信息请查看图 3-1。

3.2 数据库设计

目前常见的数据库类型主要分为两种：一种是如 MySQL、Microsoft Access、Oracle 这样的关系型数据库，另一种则是近年来逐渐兴起的非关系数据库如 MongoDB、CouchDB 等。考虑到本系统中需要储存的数据不是特别大，且无须用到复杂的查询操作，因此没有必要使用大型的数据库管理系统，再加上本系统后台使用 Express 进行开发，对 MongoDB 的支持度很高，再加上图形操作界面的支持，最终本系统选用的非关系型数据库 MongoDB 来储存系统数据。MongoDB 是一个面向文档的数据库，支持直接存取 BSON，这意味着可以在文档中直接插入复杂类型的数据，同时在使用过程时不需要事先定义关系数据库中“表”等数据库对象即可使用，节省数据库设计时间，方便项目的开发。

3.2.1 数据库集合设计

在关系型数据库中 MySQL 中，一个数据库由一个或一组数据表组成。而在非关系型数据库 MongoDB 中，一个数据库由一个或多个集合构成，这里的“结合”的概念类似于关系型数据库中的“表”，在使用过程中差异就是 MongoDB 中无须事先定义集合，当插入数据时 MongoDB 会自动判断要储存的集合是否是新的集合，若是，则会自动为用户创建，相较 MySQL 操作较为方便。本系统设计了 5 个数据集合，分别是用户信息数据集合、被测信息数据集合、试题信息数据集合、试题项目信息数据集合、测验结果信息集合。表 3-1 至表 3-5 是本系统的数据库表的详细介绍，包括各个数据库中所包含的字段名称、数据类型、主键和字段说明。

(1) 用户信息数据集合

表 3-1 用户信息数据集合

集合名: users			
说明:用来存储用户信息			
键名	数据类型	约束	说明
name	String		用户姓名
password	String		用户密码
num	String		用户学号
identity	String		用户身份
date	Date		注册时间
tests	Array		做题记录

_id	ObjectId	主键	用户编号
department	String		用户院系

用户信息数据集合包含 8 个字段，分别是用户编号、用户姓名、用户密码、用户学号、用户身份、注册时间、做题记录、以及所属院系，其中主键是用户编号，做题记录为一个对象数组。

(2) 被测信息数据集合

表 3-2 用户信息数据集合

集合名: exmaees			
说明: 存储被测个体信息			
键名	数据类型	约束	说明
_id	ObjectId	主键	被测编号
name	String		被测姓名
num	String		被测学号
department	String		被测院系
date	Date		添加时间
testId	ObjectId		试卷编号
operator	String		添加人
flag	Boolean		完成标志

被测编号为主键，被测姓名、被测学号、被测院系用于在检测被测是否有需要参加的考试并显示在首界面中，试卷编号用于查询参加相应的试卷编号值的考试的学生，完成标志用于表示学生是否完成了本次考试。

(3) 试题信息数据集合

表 3-3 试题信息数据集合

集合名: exams			
说明: 记录试题信息			
键名	数据类型	约束	说明
_id	ObjectId	主键	试卷编号
name	String		试卷名称
type	String		试卷类型
examer	String		试卷出题人
duration	String		考试时长
pending	Boolean		试卷状态
date	Date		试卷日期

remark	String		试卷备注
--------	--------	--	------

(4) 试题项目信息数据集合

表 3-4 试题信息数据集合

集合名: items			
说明: 记录试题项目信息			
键名	数据类型	约束	说明
_id	ObjectId	主键	题目编号
content	String		题目内容
source	String		题目来源
examer	String		出题人
date	Date		出题时间
options	Array		题目选项
type	String		题目类型
answer	String		题目答案
degree	Number		题目难度
division	Number		题目区分度
subordinate	ObjectId		题目从属

题目编号是主键,通过试题编号可以查询到相应信息,题目类型是一个字符串,取值为“单选题”、“不定项”、“判断题”三者之一。题目从属用于表示该测试项目属于某一份试题。

(5) 测验结果信息集合

集合名: results			
说明: 记录测试过程信息			
键名	数据类型	约束	说明
_id	ObjectId	主键	试题编号
userId	ObjectId		被测编号
itemId	ObjectId		试题的项目编号
date	Date		答题时间
answer	String		正确答案
choice	String		被测应答
flag	Number		应答结果
degree	Number		题目难度
theta	Number		被测能力值

这个集合存储被测个体中所有的答题记录，其中应答结果的数据类型为 `Number`，如为 1 表示答对，若为 0 表示答案错误。被测能力值是根据算法得出的再添加进来的。

3.3 题库的组成

3.3.1 试题库的构建

对于 CAT 而言，题库是整个系统极为重要的组成部分之一，一定程度上题库越大越好，对知识点的考察点也能更加全面，测试结果更能反映学生的掌握水平。本系统设计初衷是考查修读了计算机网络课程的学生对课程知识点的掌握能力，因此测试题目大多通过对教材的课后习题或者往年的课程试卷进行改编而来，使用此类题目更能反应学生对课程内容的掌握程度，同时也有一部分来自于网络作为题目的补充。

3.3.2 题目类型的选择

由于项目反应理论的特征函数只适合用于计算二值模型，即取值为 0 或 1，系统无法通过计算模型来判断主观题的作答结果，因此主观题在测验中不适用，使用客观题更合适。在本系统设计中，题目类型有判断题、单选题以及不定项。同时对于不定项类型的题目，只有结果完全正确才认为是正确应答，例如，正确答案为 `ABC` 的情况下，被测答案是如 `A`、`AB` 的情况不得分。

3.3.3 题目参数的确定

题目难度的取值是否合理影响着被测个体能力水平的计算，因此题目难度的设定应该力求科学、合理。本系统使用的 `Logistic` 单参数模型，只考虑测试题目难度与被测个体能力的关系，忽略猜测系数以及题目区分度的影响。在难度设定上，考虑到题目选项少通过猜测做出正确作答的情况，一般遵循判断题的难度应较低，不定项的难度应较高的原则来设定。在没有测试结果的情况下，题目的难度可以由教师给出，后续也可以根据被测群体的作答情况来进行调整。

3.4 系统关键模块的实现

3.4.1 被测个体初始能力值估计模块

由于一开始被测个体并没有接受测试，因此在考试前首先需要对被测个体的能力进行估计，之后系统在根据这个初始能力值在系统出题进行测验。本系统中采用的方式先提供一定数量中等难度的题目给被测个体，并根据被测个体的作答情况对题目的难度进行调整，在题目数量达到设定值时，根据被测个体的答题情况，估计出被测个体的初始能力值，本系统中使用的是能力计算公式为：

$$\theta = \log\left(\frac{d}{n-d}\right)$$

其中 θ 为待求得被测个体的能力值， d 为被测个体作答正确的题目数量， n 表示被测个体在本次考试中作答的题目总量。该过程代码实现如下：

（1）根据难度选题代码

```
let itemId = req.body.testInfo.itemId
let choice = req.body.testInfo.choice
//testInfo 包含 testId, itemId ,chioce,degree
Item.findOne({
  _id: itemId
}).then(item => {
  let flag = 0;
  if (item.answer === choice) {
    flag = 1
  } else {
    flag = 0;
  }
  iteminfo.flag = flag
  //保存测试情况
  new Result(testInfo).save()
  //难度范围调整
  if (flag == 1) { //作答正确，flag 值为 1，否则为 0
    left = item.degree + 0.01
    right = item.degree + 0.15
  }
  else {
    left = item.degree - 0.15
```



```

        right = item.degree - 0.01
    }
    //在题库中寻找难度范围合适的题目
    Item.find({
        testId,
        degree: { $gte: left, $lte: right }
    }).then(items => {
        //没有找到对应的题目
        if(!items) {
            res.json({ flag: 0,data:'测试结束'})
        }
        //有对象的题目，其他 items 是一个数组
        let notDone = 0
        for (let i = 0; i < items.length; i++){
            Result.findOne({
                itemId,
                userId:req.user._id
            }).then(res => {
                if (!res) {
                    res.json({ flag: 1, item: items[i] })
                    notDone = 1
                }
            })
            if (notDone == 1) {
                break;
            }
        }
    })
})

```

（2）能力计算代码

```

Result.find({ userId: req.user._id, testId: testId }).then(items => {
    let theta = 0;
    let item_count = items.length
    let count = 0 //题目答对数量
    items.forEach(item=> {
        if (item.flag == 1) {

```

```

        count++
    }
});
theta = Math.log(count / (item_count - count))
//保存结果
User.findOneAndUpdate(
    { _id: req.user._id },
    {
        $push: {
            "tests": { testId, theta }
        }
    }, { new: true } //允许返回更新后的数组
)
})

```

3.4.2 自适应选题模块

该模块的作用是根据被测个体对当前题目的做题情况，通过相关计算从题库中选出被测个体未做过的、信息量最大的题目提供给被测个体。在不满足终止条件的情况下，重复该步骤。题目信息量的计算公式如下：

$$I_i(\theta) = \frac{\exp(\theta - b_i)}{(1 + \exp(\theta - b_i))^2}$$

其中 I_i 表示第 i 个题目提供的信息量， θ 是被测当前能力值， b_i 第 i 个题目的难度。代码实现如下：

```

Item.find({ testId: testId }).then(items => {
    let itemInfo = { itemId: "", value: 0 };
    let maxInfo = { itemId: "", value: 0 };
    let first = 0
    for (let i = 0; i < items.length; i++) {
        let people = items[i].people
        if (!people.includes(req.user._id)) {
            first++
            // theta 表示被测个体的能力
            itemInfo.value = Math.exp(theta - items[i].degree) / Math.pow(1 + Math.exp(theta -
items[i].degree), 2)
            itemInfo.itemId = items[i]._id
            if (first == 1) {

```

```

        maxInfo = {
            itemId: items[i]._id,
            value: itemInfo.value
        }
    }
    if (itemInfo.value > maxInfo.value) {
        maxInfo = {
            itemId: items[i]._id,
            value: itemInfo.value
        }
    }
}
return maxInfo
})

```

3.4.3 能力参数模块

假设有 n 道试题，被测个体能力为 θ ，第 i 题的难度为 b_i ，其中 $1 < i < n$ ，用 u_i 表示被测个体对第 i 题的作答反应，则若能力为 θ 的被测个体答对第 i 题 $u_i = 1$ ，否则 $u_i = 0$ 。用 P_i 表示能力为 θ 的被测个体答对第 i 题的概率，用 Q_i 表示打错的概率，即 $Q_i = 1 - P_i$ ，因此被测个体在该测试项目的概率分布函数可表示为 $P_i^{u_i} Q_i^{1-u_i}$ ，反应矩阵 $U = [u_i]$ 的似然估计函数为

$$L = \prod_{i=1}^n P_i^{u_i} Q_i^{1-u_i}$$

由于本系统是使用的是单参数 Logistic 模型，则

$$P_i(\theta) = \frac{1}{1 + \exp[-1.7(\theta - b_i)]}$$

$$\ln L = \sum_{i=1}^n [u_i \ln P_i + (1 - u_i) \ln Q_i]$$

根据牛顿-拉夫逊迭代可得

$$\theta_{k+1} = \theta_k - \frac{g(\theta_k)}{g'(\theta_k)}, k = 0, 1, 2, 3, \dots$$

其中 $g(\theta) = \frac{\partial \ln L}{\partial \theta}$ ， $g'(\theta) = \frac{\partial^2 \ln L}{\partial \theta^2}$ 。

根据上式进行迭代，直至达到设定的规则方可停止。在本系统中，规则设定为：

$$\theta_{k+1} - \theta_k < \varepsilon$$

其中 ε 设定为 0.005。实现代码如下：

```
User.findOne({ _id: req.user._id }).then(user => {
  let tests = user.tests
  let theta = 0
  for (let i = 0; i < tests.length; i++){
    if (tests[i].testId == testId) {
      theta = tests[i].theta
      break
    }
  }
}
Result.find({ testId, userId: req.user._id }).then(items => {
  let b = [];
  let u = []
  let p = [];
  for (let i = 0; i < items.length; i++){
    b.push(item.degree)
    u.push(item.flag)
  }
  let f1 = 0
  let f2 = 0
  do {
    let theta_0 = theta
    for (let i = 0; i < items.length; i++){
      p[i]=1/(1+Math.exp(-1.7*(theta_0-b[i])))
    }
    for (let i = 0; i < items.length; i++) {
      f1 = f1+1.7*(u[i]-p[i])
      f2 = f2-1.7*p[i]*(1-p[i])
    }
    theta = theta_0-f1/f2
  } while (theta-theta_0<0.005);
  //其他代码： ....
})
})
```

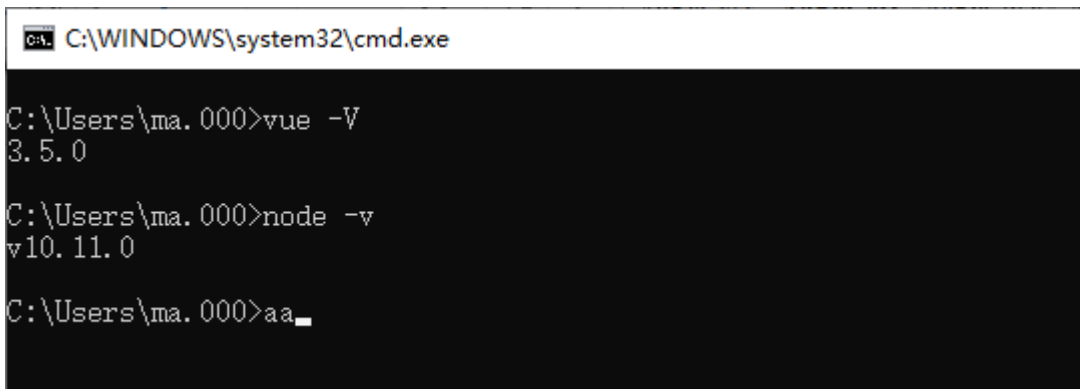
3.4 运行环境

本计算机辅助测试系统使用 JavaScript 编程语言进行开发，前端用的 Vue 框架,后端的 Node。

(1) 操作系统

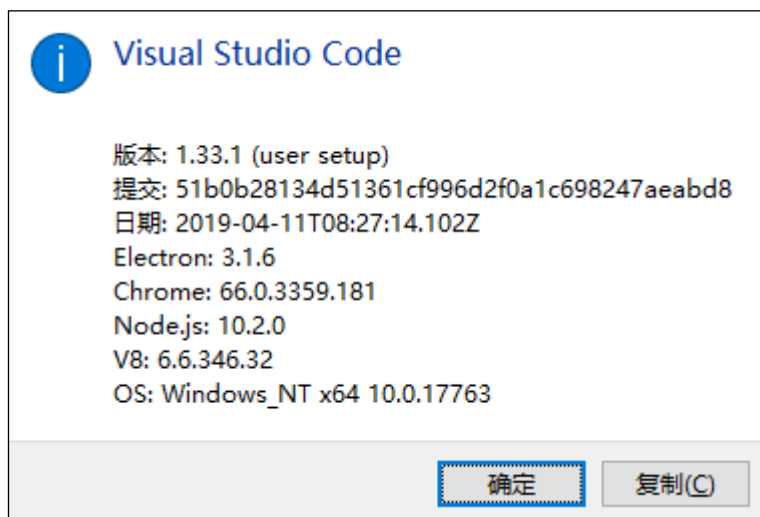


(2) Vue 以及 Node 版本

A screenshot of a Windows command prompt window. The title bar shows 'C:\WINDOWS\system32\cmd.exe'. The prompt is 'C:\Users\ma.000>'. The user has entered 'vue -V' and the output is '3.5.0'. Then the user has entered 'node -v' and the output is 'v10.11.0'. The prompt is now 'C:\Users\ma.000>aa_'.

```
C:\WINDOWS\system32\cmd.exe
C:\Users\ma.000>vue -V
3.5.0
C:\Users\ma.000>node -v
v10.11.0
C:\Users\ma.000>aa_
```

(3) Visual Studio Code



3.5 开发技术

本计算机辅助测试系统使用目前 WEB 开发中的热门框架进行开发，其中前端框架用的是 Vue.js、Element，后端框架使用的是 Express。

3.5.1 Vue 框架简介

Vue.js 是目前热门的三大前端框架之一，是一款用于构建用户界面的渐进式 MVVM 框架。使用 Vue.js 可以快速完成页面的开发。主要优点有：

- (1) Vue.js 的核心库只关注视图层，开发者可以专注于逻辑的实现
- (2) 支持双向数据绑定，无须直接操作 DOM
- (3) 学习曲线低，易于上手——只需要有良好的 HTML 和 JavaScript 基础便可以通过阅读文档进行开发
- (4) 灵活性高——Vue 官方提供了构建工具来协助你构建项目，但它并不限制你如何组织你的应用代码。
- (5) 运行性能高——由于不使用脏检查，Vue 非常容易优化。

3.5.2 Element 框架简介

Element 是一套为开发者准备的基于 Vue.js 的桌面端组件库。Element 组件封装好了常用的逻辑以及提供常见的样式，使用 Element 组件可以极大地节省开发者在页面样式的时间投入，避免长时间关注于 CSS 样式的调整，因此可以快速搭建页面，提升产品设计效率。主要优点包含以下几点：

-
- (1) 现实生活一致性：与现实生活的流程、逻辑保持一致，遵循用户习惯的语言和概念。
 - (2) 提供控制反馈：通过界面样式和交互效果让用户可以清晰的感知自己的操作。
 - (3) 效率高：界面简单直白，让用户快速识别而非回忆，减少用户记忆负担。

3.5.3 Express 框架简介

Express 是一款基于 Node.js 的开发框架，具有开放、响应速度快等特点。主要优点包含：

- (1) Express 是基于 Node.js 后台开发框架，保持着最小规模的灵活性。
- (2) 支持许多实用的、易用的中间件，方便 API 的创建。
- (3) 使用 JavaScript 即可开发，降低前端开发人员学习成本。

3.6 系统界面

第三方刻录机

- (1) 登录界面



- (2) 注册界面

用户注册

* 用户名

* 邮箱

邮箱格式不正确

* 密码

密码不能为空

* 确认密码

选择身份

注册

(3) 首页

V 自适应测试后台管理系统
 欢迎 t111

首页
试题
信息管理

时间筛选:

筛选
 共 11 条

5条/页

<
1
2
3
>

前往 页

序号	试卷名称	试题类型	创建时间	出题人	考试时长	备注	操作
1	第2章课后习题	课堂突击	🕒 2019-03-20 03:58	周老师	30 min	"	🔗 开始考试
2	第3章课后习题	单元测试	🕒 2019-03-20 03:59	周老师	30 min	"	🔗 开始考试
3	第4章课后习题	期末测试	🕒 2019-03-20 04:00	周老师	120 min	"	🔗 开始考试
4	第5章课后习题	单元测试	🕒 2019-03-20 04:00	周老师	90 min	"	🔗 开始考试
5	第10章课后习题	突击测试	🕒 2019-03-20 07:32	蔡老师	45 min	"	🔗 开始考试

(4) 试卷发布

添加试卷信息

*

 试题类型

试题类型

▼

*

 考试时长

考试时长

▼

*

 试卷名称

*

 出题人

备注:

取消

提交

(5) 考生列表页

第一章

单选题

难度1.122

marong

网桥是用于 () 的设备。

- ☐ 物理层
- ☐ 网络层
- ☐ 应用层
- ☐ 数据链路层

提交

致谢

在此，特别感谢我本次毕业设计的指导老师——周腾老师，周腾老师在本次系统的设计中给我悉心指导与帮助，对于在设计过程中不懂、不了解的问题，周老师总是牺牲自己的时间与我进行开会讨论，一对一指导，并监督完成情况，因此特别感谢周腾老师的指导，没有他，我无法顺利地完成本次毕业设计。

同时，还要感谢在本次毕业设计中，曾经给予我帮助的同学，在我陷入困境的时候，给予我帮助，给我加油打气，让我充满动力。

感谢我的父母，他们给了我很多无法衡量的东西，他们在起早贪黑地辛苦工作，供我读书，让我完成学生，希望以后自己有能力不负他们对我的期待，能够报答他们，让他们过上好日子。

参考文献:

- [1]丁金芳. 基于 Web 的化学自适应测验系统的研究[D]. 南京师范大学, 2002.
- [2]梁丽娟. 适应性语言测试的发展、问题及出路[J]. 华北电力大学学报(社会科学版), 2015(06):136-140.
- [3]颜杰群. 项目反应理论在计算机自适应题库建设中的应用[J]. 齐齐哈尔大学学报(自然科学版), 2011, 27(05):24-26.
- [4]于建芳. 大学英语六级词汇自适应测试系统的研制[D]. 山东师范大学, 2016.
- [5]单美静, 刘琴. 基于项目反应理论的组卷算法研究[J]. 软件工程, 2017, 20(12):6-8.
- [6]于海霞, 刘竞杰, 王家骐. 基于项目反应理论自适应考试系统的设计与应用[J]. 合肥学院学报(自然科学版), 2010, 20(03):44-48.