基于机器学习的英语单词智能打卡系统

摘 要

随着人工智能和机器学习技术的发展，基于机器学习的英语单词学习应用程序兴起。这些应用程序可以自动检测学习者的语音或笔迹输入，给予即时反馈与评分。本研究设计并实现了一套基于机器学习的英语单词智能打卡系统。学习者可以通过手写输入单词，系统根据相关度对语料库中类似的单词和例句进行检索，给予结果并在系统中记录学习者的学习记录。相比现有冗余的英语学习应用，本系统提供更加个性化和简便，有潜力成为英语学习领域一款有价值的人工智能产品。本研究为基于机器学习的英语学习应用提供了一种新的设计思路。未来，在系统功能和模型的不断优化下，这一基于机器学习的英语单词学习方法有望被应用于更广阔的学习场景与群体中。机器学习技术可以实现英语学习内容的自动生成与优化，这一设计思路为开发人工智能英语学习工具提供了全新的视角，有望在提高学习效率与体验的同时，进一步推动教学方法的革新。

关键词　机器学习；单词打卡；自然语言处理

A Machine Learning Based Intelligent Clock in System for English Words

**Abstract**

With the advent of artificial intelligence and machine learning the technology has seen the rise of machine learning based English word learning apps. These apps can automatically detect learners' speech or handwriting input, giving instant feedback and grading. This Research designs and implements a machine learning-based English word-intelligent punch card system.Learners can input words by handwriting, and the system searches similar words and example sentences in the corpus according to the degree of relevance, gives the results and records the learners' learning records in the system. Compared with the existing redundant English learning applications, this system provides more personalization and simplicity, and has the potential to become a valuable artificial intelligence product in the field of English learning. This study provides a new design idea for machine learning-based English learning applications. In the future, with the continuous optimization of system functions and models, this machine learning-based English word learning method is expected to be applied to a wider range of learning scenarios and groups. Machine learning technology can realize the automatic generation and optimization of English learning content. This design idea provides a new perspective for the development of artificial intelligence English learning tools. It is expected to further promote the innovation of teaching methods while improving learning efficiency and experience.

**Keywords　Machine learning, Word punch, Natural language processing**

目 录

摘要 I

Abstract II

[第1章 绪论 4](#_Toc15554)

[1.1 课题背景 4](#_Toc9641)

[1.1.1 研究意义 4](#_Toc23675)

[1.1.2 国内外研究现状分析 6](#_Toc17569)

[1.1.3 系统需求分析 7](#_Toc716)

[第2章 系统设计 11](#_Toc20265)

[2.1 系统的基本数据 11](#_Toc21860)

[2.1.1 数据设计 11](#_Toc16621)

[2.2 系统的概要设计 13](#_Toc790)

[2.2.1 功能模块 14](#_Toc28669)

[2.3 系统的详细设计 17](#_Toc2579)

[2.3.1 架构设计 17](#_Toc26787)

[2.3.2 模型设计 20](#_Toc6381)

[2.4 系统的核心算法 21](#_Toc27855)

[2.4.1 对比算法 21](#_Toc10456)

[2.4.2 评价指标 21](#_Toc23523)

[2.5 本章小结 21](#_Toc8893)

[第3章 系统编码及运行 22](#_Toc11756)

[3.1 系统开发涉及的软件 22](#_Toc11804)

[3.2 系统运行界面及结果 22](#_Toc17796)

[第4章 系统测试 23](#_Toc32724)

[4.1 测试方法 23](#_Toc13816)

[4.1.1 方案一 24](#_Toc28740)

[4.1.2 方案二 24](#_Toc18276)

[4.2 结果分析 25](#_Toc7016)

[4.3 本章小结 25](#_Toc4321)

[结论 26](#_Toc8921)

[致谢 27](#_Toc252)

[参考文献 28](#_Toc10460)

# 绪论

## 课题背景

随着人工智能和机器学习技术的发展，基于机器学习的英语学习应用程序兴起，如无人英语老师和AI英语家教等。这些应用程序可以自动检测学习者的语音或笔迹输入，给予即时反馈与评分，实现个性化和高效的学习。相比传统英语学习方法，这些新应用程序有望极大优化学习体验，激发学习兴趣，提高学习效果。但是，现有英语学习应用程序仍存在一定局限。大多系统需要人工录入大规模的单词、词性、例句与翻译，无法实现学习内容的自动生成，增加了开发工作量，难以扩展。另外，现有系统主要基于规则引擎与统计机器翻译，学习内容与效果难以做到高度个性化，无法充分满足不同学习者的需求。基于此，研究开发一套基于机器学习的英语单词智能打卡系统具有重要意义。通过对海量公开语料的采集、清洗与向量化，机器学习模型可以实现学习内容的自动生成与优化。与规则引擎和统计方法相比，机器学习模型可以建立单词语义空间，提供高质量的近义词、例句与翻译，实现学习内容的高度个性化。系统可以实现海量英语语料的下载、清洗与学习内容的自动生成，提供个性化的学习反馈与评分，优化学习体验与效果。研究成果有望为英语学习应用的开发与优化提供新思路，推动机器学习在英语教学领域的应用与发展。本课题旨在通过机器学习技术，研发一套基于机器学习的英语单词智能打卡系统。系统可以实现学习内容的自动生成与优化，提供高度个性化的学习体验，这一设计思路将为基于人工智能的英语学习工具开发开拓新视野，有望提高学习效率与学习效果，为英语教学方法的革新与发展注入新动力。

### 研究意义

许多专家学者研究表明，词汇量与语言能力及认知能力密切相关。如美国著名语言学家诺曼·莱文斯指出:“词汇是思维的工具，词汇是思想的武器。”我国语言学家桂诗春也研究发现，词汇量大的人语言水平不一定高，但语言水平高的人词汇量都比较大；词汇量测试和智力测试高度相关。由此可见，记忆词汇是学习英语最基本的要求。只有不断丰富英语词汇，听、说、读等语言能力才能提高。通过单词打卡系统，学习者可以随时随地学习记忆词汇。系统分析学习者的记忆遗忘规律，在关键时间点智能提醒，将短期记忆转化为长期记忆，从而提高学习效率。 总之，专家学者的研究结果显示，词汇量是反映语言与认知能力的关键指标。要提高英语水平，词汇量的积累至关重要。单词打卡系统可持续监测学习者的记忆情况，在关键时间点开展干预，使词汇从短期记忆进入长期记忆，这大大提高了词汇记忆的效率，对英语水平的提高起到关键作用。通过这一系统，学习者可以有效记忆词汇，为全面提高英语能力奠定基础。

本选题“基于机器学习的英语单词智能打卡系统”中的两个关键词为“机器学习”与“智能单词打卡”。

机器学习是一门融合多学科知识的交叉学科，其中涉及概率论、统计学、近似理论以及复杂算法等知识。机器学习运用计算机作为工具，致力于模拟人类学习的方式，对知识进行结构化处理，以提高学习效率。它的研究对象是如何使用计算机模拟或实现人类的学习活动，这也是人工智能领域最具智能特征和最前沿的研究方向之一。本系统在打卡任务生成、打卡评判两部分应用到机器学习、自然语言处理等技术对打卡任务进行处理。

本研究实现了基于移动端开发的系统进行英语词汇的学习，为建立以学习者为中心、打卡学习以及自我控制学习进度的方式探索出一条新途径。其次，本文通过移动端为学习者提供基于具体语境的初级词汇打卡学习，并使用词义之间的聚合将词汇进行分类，有助于学习者构建系统的词汇结构，使学习者能够达到更好地学习词汇的目的。软件推出后，学习者可以选择不同的方式学习英语词汇，这大大提高了词汇学习的方便性以及记忆效率。软件利用大数据分析，找到背诵单词的效率规律和遗忘曲线，然后根据曲线来规划记忆时间，安排单词出现的几率，确保背诵单词的最佳效率。教师端根据单词遗忘曲线程度安排考核内容，提高学生背单词效率以及教师工作效率。

#### 理论或实际应用方面的价值

（1）理论意义

随着移动互联网技术的不断发展，基于移动智能终端的移动学习逐渐被广大学习者所接受，通过基于手机端的词汇学习平台设计与实现的研究，可以进一步明确移动学习的内涵，有助于移动学习理论进一步发展与完善。

（2）实践意义

首先，本研究实现了基于移动端进行英语词汇的学习，为建立以学习者为中心、自我控制学习进度以及个性化学习的方式探索出一条新途径。其次，本文通过移动端为学习者提供基于具体语境的初级词汇学习，并使用词义之间的聚合将词汇进行分类，有助于学习者构建系统的词汇结构，使学习者能够达到更好地学习词汇的目的。最后，本系统主要针对二年级以上大学以下学生，对于中小学生高效率的提升词汇量，具有一定的现实意义与价值。

### 国内外研究现状分析

#### 国内研究现状

随着时代的进步，互联网移动网络技术的不断发展和创新，以及教育技术的改革和深化，各类移动学习方式和平台的涌现、教育培训组织、学习类、工作类、生活类软件的应用，促使技术人员不断开发、更新换代移动软件，以满足人们生活、工作和学习需要。英语作为在国际上最主要的交流语言，国内学生学习和备考的重要学科，教育的策略和学习的方式方法也受到影响，词汇的学习是学好一门语言的基石，在现代网络发达的社会，在英语词汇的学习过程中，移动学习成为一种主要的方式。在这样的社会背景下，移动知识学习的改革正在实际的教学教研活动以及相关领域中发生着。

移动学习在教育上的应用早已被教育科研人员所关注，移动学习领域的重要研究方向之一就是语言学习。早在2000年，斯坦福大学学习实验室就启动了一项关于此的重要项目。他们通过手机这种移动的方式对西班牙语的学习开展研究，结果证明，学习者通过使用手机的功能可以使得学习语言的效果明显提高。

传统的英语单词学习工具包括纸质资源，如单词手册、课本和词典等。这些工具的限制在于，学习者必须按照固定的单词顺序线性地学习，无法进行任何交互，并且无法根据学习者的学习特征进行调整和反馈。然而，在移动学习软件的广泛应用下，大量英语单词移动学习软件开始涌现，这些软件按照功能特点可分为“记忆单词”、“单词助记”和“记忆管理”三个层次。“背单词”属于最低层次，这种软件只是将纸质词典电子化；“单词助记”则利用辅助信息或媒体帮助学习者有效地记忆单词；而“记忆管理”则属于最高层次，它可以根据遗忘规律精确地安排大量单词的复习。

#### 国外研究现状

国外移动语言学习的研究以美国斯坦福大学建立了全球第一个移动语言学习实验室为标志。他们首先通过手机对西班牙语的学习进行了研究，其研究结果表明学习者借助手机的语音信箱功能与E-mail功能够明显地提高学习者的语言学习效果。随后移动语言学习的研究在世界范围内展开。在日本，Thornton和 Houser开展了一项关于词汇教学的实践性研究，研究者们通过手机短信向学习者推送有关课程的学习内容，其课程学习内容推送频率为一天三次且内容简明扼要，研究结果表明通过手机短信进行词汇学习的学习者的学习效果明显高于采用传统课堂学习的学习者。

#### 总结对比分析

随着移动互联网技术的发展与进步，目前，国外对移动语言学习的研究领域更多地注重对移动智能终端开展语言学习方面的研究，尤其是语言听力方面的学习研究更为突出，这也是世界范围内的共同研究趋势。

移动语言学习作为一种不受时间与空间限制的信息化学习方式，不仅能够为学习者制定符合自身的个性化学习方案，而且还可以将传统课堂学习活动延伸到生活当中，与此同时也能够给予学习者及时的学习反馈，对学习者学习效果的提升起到了促进作用。正是由于移动语言学习方式在语言学习方面具有诸多的优点，才被人们视作辅助语言学习的首选学习方式。随着技术的不断进步与成熟，移动语言学习与英语学习之间的联系将越来越紧密，其发展前景将会越来越广阔。

### 系统需求分析

#### 功能需求

该毕业设计是开发一个智能单词打卡系统，该系统主要有四个功能模块，分别是注册登录模块、打卡任务模块、遗忘曲线分析模块、班级权限管理模块。用户通过注册获得账号，然后进行登录，可以进行个人信息修改，系统共划分两类身份，分别是学生、教师。在打卡任务模块中主要功能有教师发布打卡任务，查看学生完成情况，班级成员权限管理。在遗忘曲线分析模块中，主要的功能有根据学生单词记忆情况分析出记忆遗忘曲线，根据曲线遗忘程度着重安排单词出现频率，最大程度的提升背单词效率以及单词的记忆程度，教师根据打卡任务完成情况，单词错误率安排个别单词考察，提升工作效率。班级权限管理模块中，学生可以搜索班级号加入班级，退出班级操作。教师可以将学生加入到自己的班级，也可以将学生移出班级。例如教师注册账户后可以创建多个班级，每个班级有唯一的班级号，学生可以搜索班级号加入班级，也可以通过教师添加的方式进入班级。根据需求，制作系统功能模块划分结构图，如图1-1所示：



图 1-1系统功能模块划分结构图

各个模块功能如下：

1. 注册登录模块：

用户通过注册获得账号，通过账号进行登录使用系统，用户登录后要进行身份认证，通过账号密码等信息验证学生或老师的身份，不同的身份所展示的效果不同，不同的身份都可以进行个人信息修改。

1. 打卡任务管理：

以教师身份登录单词打卡系统后，可以选择班级发放打卡任务，以学生的身份登录单词打卡系统后，可以对老师发放的单词打卡任务进行完成，任务完成后后台自动评判产生成绩。

1. 遗忘曲线分析模块：

以教师身份登录单词打卡系统后，可以学生的班级权限管理，根据学生单词完成情况，重点考察错误率较高，易错的单词，以学生的身份登录单词打卡系统后，单词每隔一段时间会再一次出现，根据完成情况系统分析出遗忘曲线，根据遗忘曲线安排后续单词出现的频率，最大程度上保证背单词的效率以及成功率。

1. 班级权限管理模块：

在这个模块中，参与者是老师和学生，教师可以创建班级，添加学生，移出学生。学生可以选择加入班级，，退出班级。事件会同步在一个关系内的用户群体中。

#### 非功能需求

1. 性能需求：用户期望该软件在响应速度、结果精度和资源消耗等方面表现良好。
2. 可靠性需求：用户要求该软件尽可能不失效，并要求易恢复性，系统故障易预测且严重程度尽量降低。
3. 易用性需求：该软件需要具备易理解、易学习、易操作等特点，人机界面应简单明了，用户无需手册即可顺利进行各种操作。
4. 运行环境约束：用户对软件运行的环境有特殊要求。
5. 外部接口：用户要求该软件与其他软件系统或硬件设备具有互通性。
6. 安全性需求：该软件需要保证用户的重要信息安全，不同用户只能查看自己相关信息，具有最高权限的管理员才可修改系统，系统需提供数据备份及意外崩溃时数据恢复等功能。
7. 可保障性需求：用户要求该软件具备可配置、可扩展、可维护、可移植等特点。

#### 需求模型

1) 学生具备的功能有：登录、注册、个人信息修改、完成打卡任务、加入班级，退出班级。

2) 老师具备的功能有：登录、注册、个人信息修改、发布打卡任务、查看打卡任务完成情况，创建班级，拉取学生，移除学生。

学生老师用例图如图1-2所示：

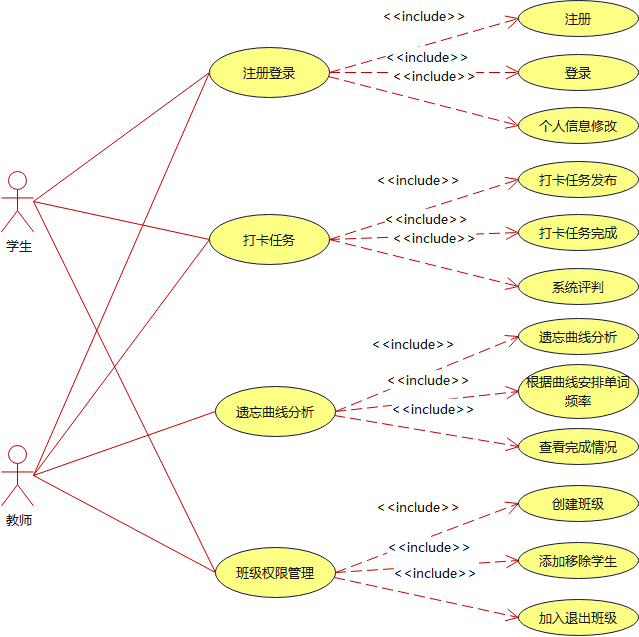


图1-2学生老师用例图

3）系统用例图如图1.3所示：

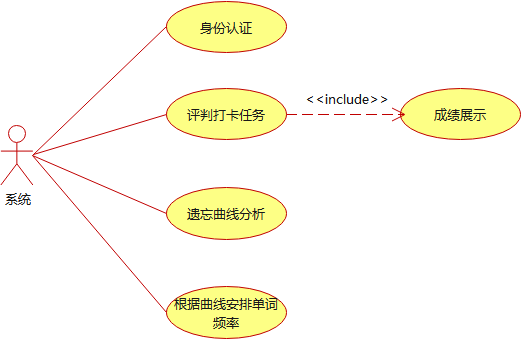


图1-3系统用例图

#### 可行性分析

1. 技术可行性：机器学习方法已经广泛用于自然语言处理，其在单词分类、预测和翻译方面的表现已经超越了人类。因此，我们可以使用机器学习算法来实现基于英语单词的智能打卡系统。
2. 数据可行性：目前，大量的英语单词语料库和单词库已经公开发布，可以使用这些数据来训练和优化机器学习模型，同时也可以通过数据挖掘和爬虫等技术来扩充和更新语料库和单词库。
3. 实施可行性：基于机器学习的英语单词智能打卡系统可以为学生、教师和英语学习者提供方便、快捷、高效的学习工具，具有很高的市场需求和潜在用户群体。
4. 安全可行性：在系统的开发过程中，可以使用各种安全技术和措施来保护用户数据的安全和隐私，例如数据加密、用户认证、访问控制等，以确保系统的安全性和可靠性。
5. 综上所述，基于机器学习的英语单词智能打卡系统技术方案具有很高的可行性，可以为用户提供优质的学习体验和商业价值，同时也可以促进机器学习技术在教育领域的应用和发展。

# 系统设计

## 系统的基本数据

本系统的数据采用关系型数据库mysql 进行数据的存储。存储的信息包括各种事务的 id（包括分享事务，关注 id 以及用户的全局 id），用来唯一标识个体及其事务、按照用户 id 索引的用户信息（包括用户注册邮箱、 用户名以及用户密码）、按照用户名进行索引的用户信息（包括用户 id）、用户存储的记事本（包括用户 id 以及用户的记事本内容）、用户关注关系（包括记录关注的用户信息）、用户被关注关系（包括记录该用户被那些用户关注的 id）、用户分享记录（包括用户分享了哪些文件给哪些用户）。将上述存储的信息设计为 hbase 中的表格，如下进行详细说明：gid 表主要用来保存各种事务和个体的 id，用来唯一识别。这里的表分别存储记事本的 id，用户的 id，以及分享事务的 id，对应只在同名的列限定符上取值，所有的列限定符都在 gid 的列族之下。

### 数据设计

该数据库包含以下表格：

1. Users表: 该表存储系统的用户信息，包括用户ID、用户名、密码。

id\_user 表主要以用户的 id 为索引的行键，列限定符包括 email、name、

pwd，都在 user 的列族之下，主要用来按用户 id 索引信息使用。

表3-1 用户信息表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型 | 字段说明 | 是否主键 | 是否为空 |
| User\_id | Numberic | 用户\_id | Yes | No |
| User\_name | Varchar2(20) | 用户姓名 |  | No |
| password | char(7) | 密码 |  | No |
| avatar | Varchar2(30) | 头像 |  |  |
| class | numberic | 班级 |  |  |

1. Words表: 该表存储单词的信息，包括单词ID、单词名称和单词的定义等。

表3-2 单词信息表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型 | 字段说明 | 是否主键 | 是否为空 |
| Word\_id | Numberic | 单词\_id | Yes | No |
| Word\_name | Varchar2(20) | 单词名称 |  | No |
| Word\_defined | char(20) | 定义 |  | No |
| Word\_phonetic | Varchar2(30) | 音标 |  |  |

1. Word\_Usages:表：该表存储单词的用法信息，包括用法ID、单词ID、用法的语境和示例句子等。

表3-3 单词用法表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型 | 字段说明 | 是否主键 | 是否为空 |
| Usages\_id | Numberic | 用法\_id | Yes | No |
| Word\_id | Numberic | 用户姓名 |  | No |
| Usages | Varchar2(30) | 用法 |  | No |
| example | Varchar2(30) | 例句 |  |  |

1. User\_Words表: 该表存储用户学习单词的记录，包括用户ID、单词ID、学习时间和学习次数等。

表3-4 学习记录表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型 | 字段说明 | 是否主键 | 是否为空 |
| User\_id | Numberic | 用户\_id | Yes | No |
| User\_id | Numberic | 单词\_id |  | No |
| Time | Numberic | 学习时间 |  | No |
| frequency | Numberic | 学习次数 |  |  |

1. User\_Word\_Usages表: 该表存储用户学习单词用法的记录，包括用户ID、用法ID、学习时间和学习次数等。

表3-5 学习用法表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型 | 字段说明 | 是否主键 | 是否为空 |
| User\_id | Numberic | 用户\_id | Yes | No |
| Usages\_id | Numberic | 单词\_id |  | No |
| Time | Numberic | 学习时间 |  | No |
| frequency | Numberic | 学习次数 |  |  |

1. User\_Word\_Checkins表: 该表存储用户打卡记录，包括用户ID、单词ID、打卡时间和打卡结果等。

表3-6 打卡记录表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型 | 字段说明 | 是否主键 | 是否为空 |
| User\_id | Numberic | 用户\_id | Yes | No |
| Word\_id | Numberic | 单词\_id |  | No |
| Create\_time | Numberic | 打卡时间 |  | No |
| result | Varchar2(30) | 打卡结果 |  |  |
| Correct\_rate | numberic | 正确率 |  |  |

1. Word\_Vectors表: 该表存储单词的向量表示，用于机器学习模型训练和推断。

表3-7 模型表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型 | 字段说明 | 是否主键 | 是否为空 |
| Model | Varchar2(30) | 模型 | Yes | No |
| Judge | numberic | 模型推断 |  | No |
| time | char(7) | 时间 |  | No |
| frequency | numberic | 次数 |  |  |
| architecture | numberic | 模型结构 |  |  |

1. Model\_Weights表: 该表存储机器学习模型的参数，用于模型训练和更新。

这些表共同支持基于机器学习的英语单词智能打卡系统的数据管理和分析功能，从而帮助用户更好地学习英语单词。

表3-8 参数表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型 | 字段说明 | 是否主键 | 是否为空 |
| Parameter | Numberic | 参数系数 | Yes | No |
| Parameter\_x | numberic | 参数\_x |  | No |
| Parameter\_y | numberic | 参数\_y |  | No |
| Parameter\_z | numberic | 参数\_z |  |  |

## 系统的概要设计

### 功能模块

通过建设基于机器学习的英语单词智能打卡系统，为学生和教师提供一站式学习平台，掌握学生学习情况，保障学习的主动性的效率，便于教学信息的相互交流传输，实现以下目标：

通过设计单词打卡系统架构，形成基于教师学生双端学习平台，实现学习任务的有效传达与效率处理。基于机器学习的英语单词智能打卡系统分为教师端和学生端。教师端主要负责班级权限管理以及学习任务的发放。学生端主要功能为完成教师发布的学习任务。

师生访问使用单词打卡服务系统主要通过网页浏览器访问模式，方便快捷。为了方便用户使用，我们将所有数据存储在一个数据库中，并提供了一站式、安全可信、易于使用的个人账户和密码登录方式，以实现随时随地访问数据。学习平台。

#### 设计目标

任何一个为了确保系统的高效运作，系统建设必须遵守多项设计原则，包括安全性原则、系统化原则、实用性原则、稳定性原则和可维护性原则等等。在本项目中，根据具体的设计需求，我们提出了以下方案：

1. 提高学习效率：通过机器学习算法的应用，实现对学习者的单词掌握程度进行智能分析，从而提高学习效率。
2. 提高系统的智能化程度：通过采用机器学习算法，对学习者的学习情况进行全面分析，从而提高系统的智能化程度。
3. 提供个性化的学习方案：根据学习者的学习情况和单词记忆情况，提供个性化的学习方案和建议，从而更好地帮助学习者掌握英语单词。
4. 提供多种学习方式：除了传统的单词记忆方式外，系统还应提供多种学习方式，例如单词拼写、听力理解等，满足不同学习者的需求。
5. 实现可扩展性：系统架构应该具有可扩展性，能够方便地添加新的学习功能和模块，以适应不断变化的学习需求。
6. 提供良好的用户体验：系统应该提供良好的用户界面和交互体验，方便用户使用，从而提高学习积极性。

#### 系统功能

本系统的研发目的在于提高英语学习者单词记忆的效率与效果，帮助用户养成规律而持续的背诵习惯。为实现此目标，系统将采取积极主动的策略来推动用户学习，这一理念贯穿于系统的每一功能设计之中。基于此设计理念，系统的具体功能不会是简单被动的知识展示，而是要主动监测用户的学习情况，并在需要时进行针对性地提醒与推动。系统首先需要建立用户的学习模型，追踪用户的学习热度与关注词汇，预测用户今后一段时间内的学习需求。其次，系统要根据用户模型定期推送合理的学习计划给用户，并在学习过程中随时给出提醒， 维持学习的连续性与强度。再次，系统还会根据用户的实际学习表现不断修正用户模型，实现模型的动态优化，以保证提醒与推送的准确性。系统的主要功能如下：

1. 用户管理功能：该功能提供用户注册、登录和个人信息管理等功能，以便用户可以使用系统并查看自己的学习记录。

用户注册和登录模块主要实现用户的登录、注册。登录需要输入用户名和密码，如果信息输入正确则跳转到主页，如果信息输入错误则跳转到登录页面；注册需要填写用户信息，如果信息输入正确，则返回登录页面，如果错误则提示注册失败。用户登陆注册流程图如图2-3 ，2-4所示：

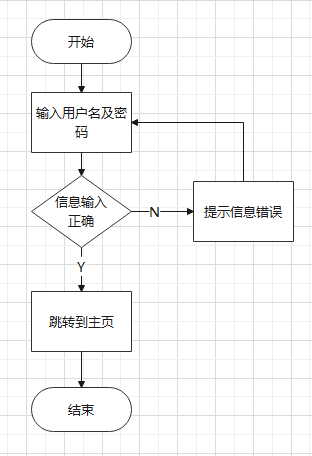


图2-3 用户登录流程图

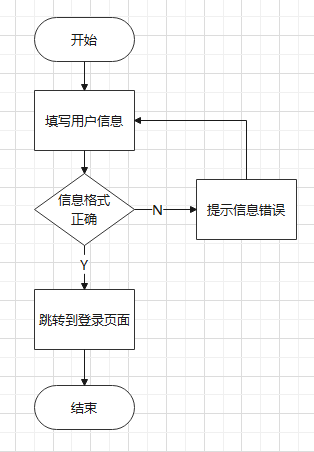


图2-4 用户注册流程图

1. 单词输入功能：该功能允许用户输入要学习或使用的英语单词，并将其传递给单词检测模块进行检测。
2. 单词检测功能：该功能使用机器学习模型和规则引擎检测用户输入的英语单词是否被正确使用，并提供错误提示和纠正建议。
3. 单词定义功能：该功能允许用户查看英语单词的定义、用法和语境，以便用户更好地理解和使用单词。
4. 学习记录和统计功能：该功能记录用户输入的单词和检测结果，并提供单词学习历史记录和统计信息，例如单词学习次数、错误率等。
5. 数据收集和预处理功能：该功能负责从不同来源收集英语单词数据，并将其进行预处理，例如清理数据、去除停用词、进行词干提取、词向量化等。
6. 模型训练和更新功能：该功能使用机器学习算法对预处理后的数据进行训练，以学习单词的定义、用法和语境，并不断更新和优化模型。
7. 数据存储功能：该功能负责存储用户输入的单词、模型训练的数据和模型参数等重要数据。

这个基于机器学习的英语单词智能打卡系统的功能结构设计可以帮助用户学习英语单词，并提供错误提示和纠正建议，从而提高英语写作和口语的准确性和流畅性。同时，它可以为学习者提供更好的学习记录和统计信息，以便他们更好地掌握自己的学习进度和成果。

#### 系统技术架构

本系统的目的在于通过机器学习技术为英语学习者提供良好的学习计划，全面提高单词记忆的效率与效果。为实现此目的，系统必须具备动态理解每个用户的学习情况并进行主动推动的能力。这就要求系统基于海量学习数据构建用户模型，并持续优化这些模型以追踪用户的最新学习状态。

系统技术架构设计图通常包括以下组件：

用户界面：提供给用户进行单词打卡、学习进度追踪、单词查询等功能的图形用户界面（GUI）。

数据库：存储单词数据、用户信息、学习记录等数据的数据库。

后端服务器：提供基于机器学习的单词学习算法，包括单词推荐、学习计划生成等功能的服务器端应用。

前端服务器：处理用户请求，与后端服务器进行交互并返回数据给用户的服务器端应用。

机器学习模块：使用机器学习算法进行单词推荐、学习计划生成等功能的模块。

系统技术架构设计图如图2-2所示：

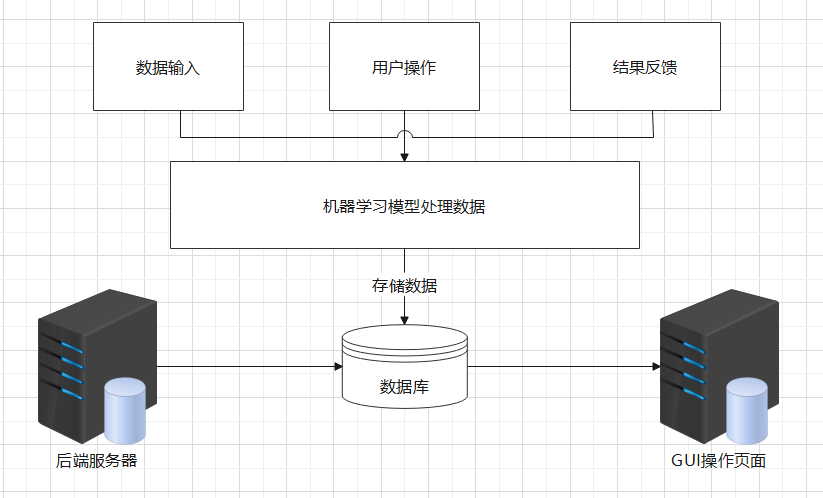


图2-2 系统技术架构

## 系统的详细设计

### 架构设计

#### 总体架构

系统总体架构自顶向下主要包括４个层次，包括：

1. 数据层：该层负责管理数据的存储和处理，包括单词库、用户数据等。
2. 处理层：该层负责对用户的操作进行处理，包括用户的登录、注册、单词打卡等功能的处理。
3. 机器学习层：该层负责训练和应用机器学习模型，对用户进行单词学习进度的预测和推荐。
4. 应用层：该层是整个系统的核心层，负责将机器学习层的推荐结果反馈给用户，并提供用户操作的界面。

系统总体架构图如图2-1所示：

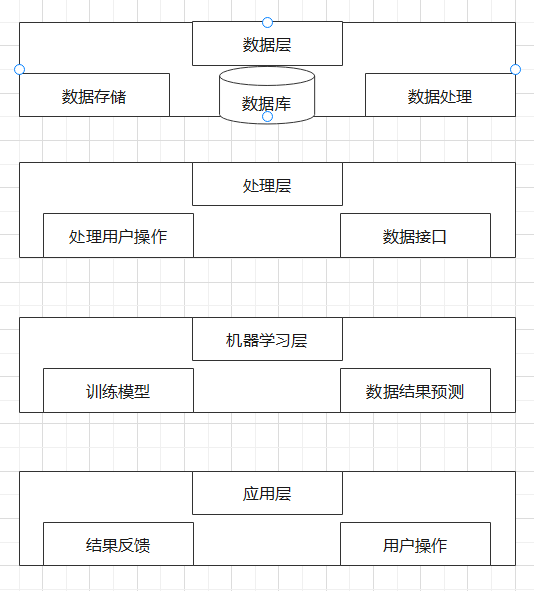


图2-1系统总体架构

#### 功能结构设计

系统工作流程为主要如下，系统分为教师端学生端，教师学生进行登陆注册，教师创建班级并发布打卡任务，学生选择加入班级，完成每日教师发布的打卡任务，系统根据学生背单词情况好坏安排后续的单词出现频率，教师端还能看到学生的单词完成情况，以及错误率，以便安排后续的考察。系统工作图如图2-5所示：

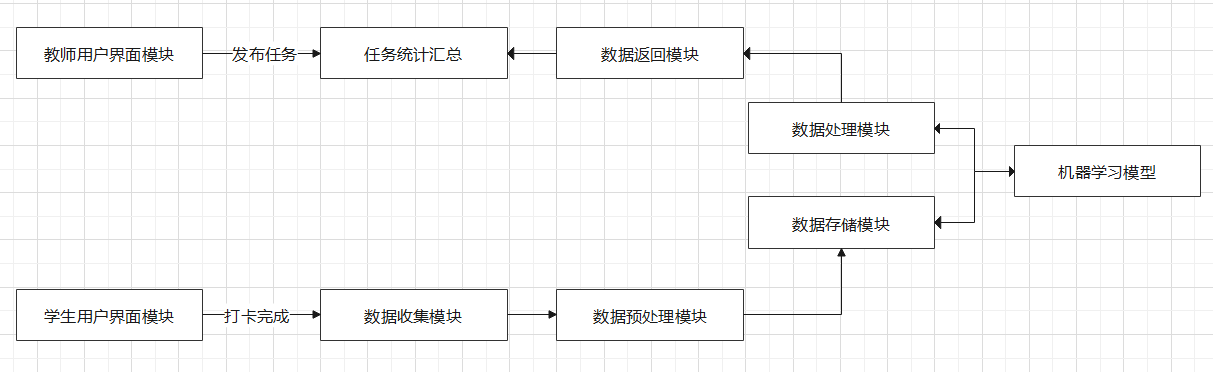


图2-5 系统工作流程

1. 数据收集模块：该模块负责从不同来源收集英语单词数据，并将其存储到数据库中。可以使用公共数据集，或从用户生成的数据集中获取数据。
2. 数据预处理模块：该模块负责对收集的英语单词数据进行预处理，例如清理数据、去除停用词、进行词干提取、词向量化等。这些预处理步骤可以提高后续模型训练的效果。
3. 模型训练模块：该模块负责使用神经网络和支持向量机等机器学习算法，对预处理后的数据进行训练，以学习单词的定义、用法和语境。可以使用已有的预训练模型，例如GPT-2、BERT等。
4. 单词检测模块：该模块负责接收用户输入的英语单词，并使用训练好的模型检测单词是否被正确使用，并提供错误提示和纠正建议。可以使用模型预测或规则引擎进行单词检测。
5. 用户界面模块：该模块负责提供用户界面，以便用户输入单词和查看错误提示和纠正建议。用户界面可以是Web应用程序、移动应用程序。具体的页面展示流程图如下图2-6所示：

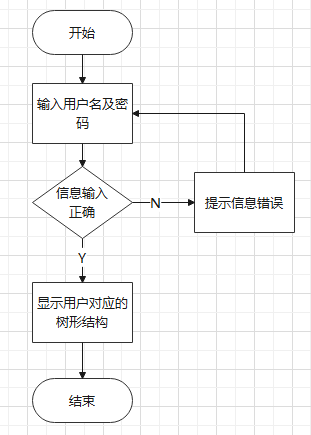


图2-6 页面展示流程图

用户对于系统的操作就是发布打卡任务与打卡任务的完成，其他的操作也同理，用以下的数据流图2-7来表示。

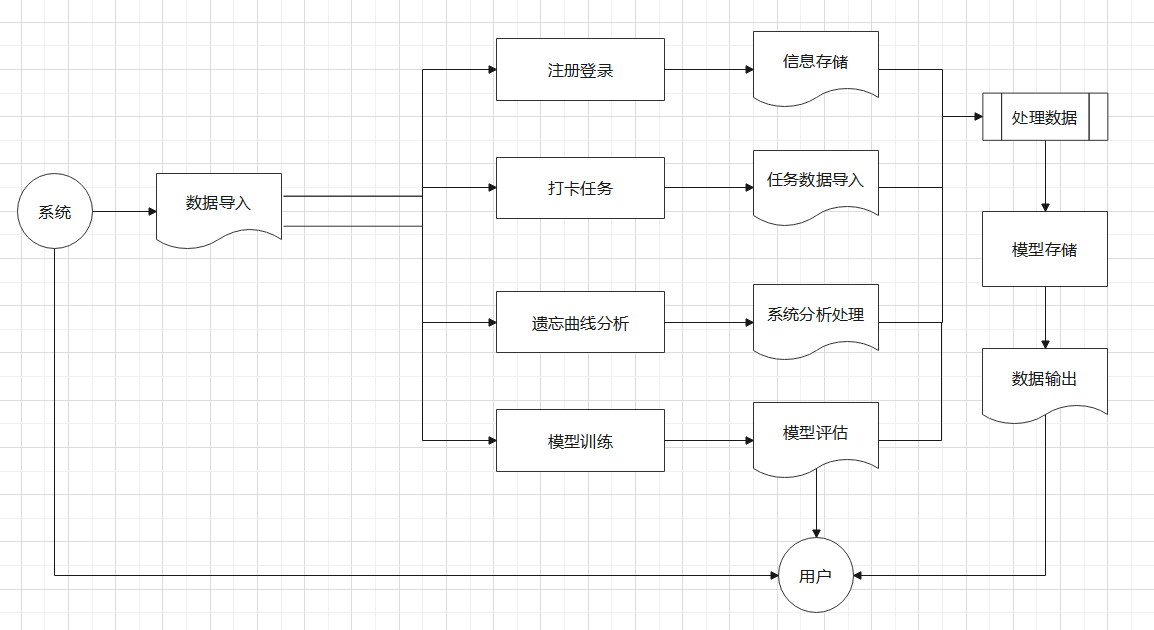


图2-7 文件操作数据流图

1. 数据存储模块：该模块负责存储用户输入的单词、模型训练的数据和模型参数等重要数据。可以使用关系型数据库或NoSQL数据库进行数据存储。

总体而言，这个基于机器学习的英语单词智能打卡系统的架构设计可以帮助用户学习英语单词，并提供错误提示和纠正建议，从而提高英语写作和口语的准确性和流畅性。

### 模型设计

#### 特征提取

#### 用户模型

#### 误差模型

#### 训练流程

## 系统的核心算法

如图

### 对比算法

所以漏磁

### 评价指标

所以漏磁

## 本章小结

本章主要建立了……，给出了……，对……进行了推导……

# 系统编码及运行

## 系统开发涉及的软件

表3-1开发软件

|  |  |
| --- | --- |
| 操作系统 | windows11 |
| 数据库 | Mysql关系数据库 |
| 使用语言 | Python、Vue.js、Python |
| 使用框架 | ElementUI |
| 开发软件 | Webstorm、Pycharm、Navicat、Postman |

## 系统运行界面及结果

# 系统测试

测试是判断一个人工智能系统质量与价值的关键手段，它使理论与技术框架实际转化为现实世界中的功能与创造价值。对于本系统来说，测试的目的在于：

第一，深入验证其设计理念的科学性与可行性。系统基于主动监测与优化学习过程这一设计理念来实现提高学习效果的目标，测试可以判断这一理念在系统实现过程中是否达成预期，如果未达成则可以发现理念实施过程中存在的失误或限制。这使理念得以在实践中进一步优化与改进，是实现理论创新与技术创造契合的关键步骤。

第二，全面评估系统技术框架的适用性与完整性。系统采用机器学习和推理技术构建自动提醒与内容推荐机制，测试可以检验这一技术框架是否能够胜任实际学习场景的需求，如果不能则可以发现其优缺点与不足之处，为技术框架的进一步优化提供依据。这使技术框架能够与实践场景完美契合，真正发挥其功效。

第三，判断系统在用户体验等方面与目标用户的契合度。作为一款教育产品，提供出色的用户体验对系统至关重要。测试可以直接反映目标用户使用系统的体验，发现系统在交互、界面以及个性化体验等方面可能存在的失误或不足之处，为下一步提高用户体验提供改进方向。这使系统的功能转化为真实的价值，并受到用户的青睐与采纳。

第四，评估系统对学习效果的提高，这是判断一个教育系统成效的关键指标。

测试提供全面深入的评价。验证系统设计理念和技术框架的适用性，发现系统问题，为改进提供依据，衡量系统成效的关键手段。良好测试判断系统质量与特点，使理念框架系统功能转化为真实世界中创建价值。测试使产品由设想转入实践，让理论联系实际，是人工智能系统顺利实施的重要环节。

## 测试方法

1. 交叉验证:这是一种统计方法，通过将数据分成训练集和测试集，训练多个模型，并在测试集上评估其表现来判断学习算法的效果。交叉验证可以有效避免过拟合，评估模型的泛化能力。本系统可以采用k折交叉验证，将数据集分成k个子集，每次选取其中一个子集作为测试集，余下的k-1个子集作为训练集，用训练集训练模型，在测试集上评估其表现。反复进行k次操作，每次使用不同的子集作为测试集，最终得出平均误差作为模型的性能度量。重复k次，最终综合k次的测试结果评估学习算法。
2. precision-recall 分析:这是一种用于信息检索系统的测试方法。Precision 表示返回结果中正确的比例，recall表示应返回的正确结果的比例。通过计算不同参数下的precision值和recall 值，可以绘制出 precision-recall 曲线，评估系统的性能。这可以用于本系统中单词和例句的推荐测试。
3. ROC分析:这也是一种用于信息检索系统的测试方法。ROC是 receiver operating characteristic 的缩写，用于判断模型分类的优良程度。通过计算不同分类阈值下的真正率和假正率，可以绘制ROC曲线。ROC曲线下的面积(AUC)可以作为模型分类效果的量化指标，值越接近1表示分类效果越好。这同样适用于本系统的推荐测试。
4. 离线评估:这种测试方法利用历史数据建立系统，然后利用未来的真实数据来测试系统效果。这需要运用到时间序列的前后拆分，一部分时间序列用于训练，另一部分用于测试。这可以避免过拟合，在真实环境中评估系统泛化能力。本系统可以利用历史学习行为数据建模，然后利用新产生的数据进行效果评价。
5. 在线评估:与离线评估相比，在线评估是在系统实际运行过程中进行的，能够反映系统在真实复杂环境中的表现。在线评估需要采用 A/B testing 等方法，同时运行不同系统，对比其效果。本系统部署后，可以采用在线评估来判断系统在实际学习场景中的影响。

### 方案一

#### 样例数据

#### 测试结果

### 方案二

#### 样例数据

#### 测试结果

## 结果分析

1）数据采集问题：为了训练机器学习模型和构建单词库，系统需要大量的语料库和相关数据。这些数据可能很难获取，特别是对于一些不常见的单词和语言结构。

解决方案：可以使用一些已有的英语单词数据集，并将其与网上公开的语料库结合使用。此外，还可以通过对用户学习数据进行分析来识别有用的单词和语言结构。

2）人机交互问题：由于系统需要和用户进行交互，可能存在用户体验不佳的问题，例如操作复杂、页面设计不合理等。

解决方案：可以进行用户体验测试和调查，以确定用户的需求和偏好，并优化系统的界面设计和功能交互。

3）系统性能问题：由于系统需要处理大量的单词和学习数据，可能会导致性能问题，例如响应时间过长或资源占用过多。

解决方案：可以优化系统的代码和算法，增加硬件资源以提高系统性能，并使用缓存等技术来减少对后端数据库的访问次数。

综上所述，这些问题并不是无法解决的，只要在系统设计和开发过程中充分考虑这些问题，并采取合适的解决方案，就可以开发出高效、准确和用户友好的基于机器学习的英语单词智能打卡系统。

## 本章小结

结论

回顾本系统的研发过程，感到欣慰的是其设计理念和技术框架初步达成了预期目标。通过主动监测和优化学习过程，系统可以为不同学习者定制个性化的学习体验，从而提高学习效率与效果。这充分体现了人工智能技术在教育领域的应用潜力，也为相关研究提供了一个可供借鉴的案例。

在设计理念上，提出通过积极主动的策略来推动学习过程，实现学习记忆的自动化与持续化。这一理念被系统的每一功能所具体体现，并且初步达到了设计目标。技术实现上，采用数据驱动和机器学习技术构建了一套可用于优化学习记忆的系统框架。这一框架能够动态理解学习者的需求，并提供定制化支持，从而实现理念的落地。

当然，也意识到本系统还存在较大提升空间。首先，系统规模有限，学习行为数据和词汇知识库有待大幅扩充，这将进一步提高系统推荐的准确性与广度。其次，我们仅采用简单的机器学习模型，更为高级的模型如神经网络有望进一步精进用户模型，增强系统的学习能力。再次，仅在小范围内验证了系统对学习效果的提高，更加广泛的评估有助于进一步检验设计理念的可行性。最后，有待深入探索如何为不同学习者和学习阶段提供最为贴心的学习支撑，这需要理论与技术的更加深入融合。

展望未来，将致力于进一步完善本系统，以实现设计理念的全面落地。首先，希望在更大范围内收集学习行为数据与构建词汇知识库，这将极大拓展系统的应用范围。其次，将探索使用更为高级的深度学习模型来构建用户模型，进一步提高系统的个性化水平。再次，希望对更广泛的学习群体进行系统评估，检验设计理念在不同情景下的适用性。最后，将不断完善如何为多样学习者提供定制学习方案，这需要进一步融合学习理论与人工智能技术。

总之，本项目启发我们进一步开拓人工智能在教育领域的研究与应用。提出的理念与框架为学习过程提供优化支持，这也为我未来在相关领域的深入工作奠定了基础。我们相信，人工智能的发展将深刻重塑教育模式，而我们的努力也将为此贡献一份力量。让我们携手努力，推动教育变革，创造更加美好的未来!

致谢

感谢我自己！

参考文献

1. 戈宝军，梁艳萍，温嘉斌．电机学[M]．第二版．北京：中国电力出版社，2013：1-3．
2. 主要责任者．题名：其他题名信息[文献类型标志]．其他责任者．版本项．出版地：出版者，出版年：引文页码
3. 余敏．出版集团研究[M]．北京：中国书籍出版社，2001：179-193．
4. 昂温G，昂温P S．外国出版史[M]．陈生铮，译．北京：中国书籍出版社，1988：89-102．
5. 全国文献工作标准化技术委员会第七分委员会．GB/T 5795-1986 中国标准书号[S]．北京：中国标准出版社，1986．
6. 作者.文章名.学术刊物名，年，卷（期）：引用部分起止页
7. 李晓东，张庆红，叶瑾琳．气候学研究的若干理论问题[J]．北京大学学报：自然科学版，1999，35（1）：101-106．
8. 叶志军，游佰强，J. Rosendahl，等．基于Flux 2D 的大型同步发电机在额定运行下定子绕组匝间短路故障特性研究[J]．中国电机工程学报，2013，33(3)：125-132．
9. LCASO A N，MARQUES CARDOSO A J．Remedial Operating Strategies for A 12-Pulse LCI Drive System [J]．IEEE Transactions on Industrial Electronics，2008，55(5)：2133-2139．
10. DE ARAUJO RIBEIRO R L，JACOBINA C B，DA SILVA E R C，et al．Fault Detection of Open-Switch Damage in Voltage-Fed PWM Motor Drive Systems [J]．IEEE Transactions on Power Electronics，2003，18(2)：587-593．
11. GAMEIRO N S，CARDOSO A J M．A New Method for Power Converter Fault Diagnosis in SRM Drives [J]．IEEE Transactions on Industry Applications，2012，48(2): 653-662．
12. 作者.文章名.编者名.会议名称，会议地址，年份：引用部分起止页
13. ALCASO A N，CARDOSO A J M．Power Supply Harmonic Filter Behavior in A Twelve-Pulse LCI Drive System Under Power Converter Faults[C]．Proceedings of the Power Electronics Specialists Conference， PESC '05 IEEE 36th，Recife，16-16 June，2005：2893-2897．
14. 研究生名.学位论文题目.学校及学位论文级别，答辩年份：引用部分起止页
15. 赵相宾．基于静止变频调速系统的抽水蓄能机组起动研究[D]．天津：天津大学，2007：13-27．
16. 张侃．同步电动机软起动装置的研究[D]．哈尔滨：哈尔滨工业大学，2008：8-21．
17. 张志祥．间断动力系统的随机扰动及其在守恒律方程中的应用[D]．北京：北京大学数学学院，1998：3-15.
18. CALM S R B. Infrared spectroscopic studies on solid oxygen [D]. Berkeley: Univ. of California. 1965：1-5.
19. 刘加林．多功能一次性压舌板：中国，92214985．2[P]．1993-04-14．
20. 河北绿洲生态环境科技有限公司．一种荒漠化地区生态植被综合培育种植方法：中国，01129210.5[P/OL]. 2001-10-24[2002-05- 28]. http:/ /211.152.9.47/sipoasp/zlijs/hyjs-yx-new.asp? recid=01129210. 5&leixin.