**设计/实践报告撰写说明**

1、请先完善封面内容，专业请填写全称，如「机器人工程」。班级以年级+专业+班级填写，如「2022级机器人工程1班」。指导教师请以学校教务系统（不是学习通）上显示为准。

2、注意区分封面的课程名称，「人工智能系统设计（三）」与「人工智能系统设计实践（三）」分别为两门课程，其中设计着重于因课程任务的需要对机器人的设计要求及各自组别的机械、电控、视觉方案设计，实践着重于方案的实现、各组别的联合调试以及问题记录等。以下内容可参考，每位同学应根据自己实际学习内容按逻辑撰写，内容形式不唯一，要求至少包含三个章节及总结，每章标题应尽可能简短精炼。

|  |  |  |
| --- | --- | --- |
| 人工智能系统设计 | | |
| 章节 | 参考内容（非章节标题） | 说明 |
| 第一章 | 课程任务介绍 | 课程任务简要介绍 |
| 第二章 | XXXX的方案制定 | 整体方案 |
| 第三章 | XXXX的方案设计\相关原理介绍 |  |
| … | … | … |
| 总结 | 设计心得\总结 | 在完成设计课程过程中遇到哪些问题，如何解决，收获了哪些学习方法等，自由发挥。 |
| 人工智能系统设计实践 | | |
| 章节 | 参考内容（非章节标题） | 说明 |
| 第一章 | 算法介绍及实现 | 所用的参数和性能等 |
| 第二章 | 软件调试 | 撰写各自制作/调试的过程，记录遇到的问题等 |
| 第三章 | 系统测试、实验结果 | 调试过程，调试结果，调试问题分析等 |
| … | … | … |
| 总结 | 实践心得\总结 | 在完成实践课程过程中遇到哪些问题，如何解决，收获了哪些学习方法等，自由发挥。 |

3、打印报告时建议先导出为PDF格式，仔细检查①目录是否出现未定义书签②正文页码是否缺失③图片是否错位等格式问题，确认无误后再打印。打印要求：**A4黑白打印，封面、评分表、任务书、摘要、目录均为单面打印，正文开始采用双面打印**。



**机器人工程学院**

《人工智能系统设计实践（四）》

课程报告

2024 - 2025 学年第 二 学期

|  |  |
| --- | --- |
| **题 目** | （按所选题目填写） |
| **专 业** | 人工智能无人系统 |
| **班 级** | 2023人工智能2班 |
| **学生姓名** | 郑康杰 |
| **学生学号** | 202310179001 |
| **指导教师** | 按教务系统 |
| **日 期** | 2025年1月3日 |

**评分表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 平时成绩评分表 | | |  | 期末成绩评分表 | | |
| 类型 | 评分标准 | 成绩 |  | 类型 | 评分标准 | 成绩 |
| 学习态度  （满分30） | 项目包括：  1.学习主动性积极性；  2.签到考勤；  3.学习过程表现等。 |  | 综合能力  （满分20） | 项目包括：  1.文献检索及综合能力；  2.灵活运用知识能力；  3.设计或实践能力；  4.数据处理能力等。 |  |
| 学习进度  （满分40） | 项目包括：  1.是否按时完成课程任务；  2.课程学习节点完成情况；  3.课程任务时间规划等。 |  | 报告/说明书成绩  （满分30） | 项目包括：  1.报告/说明书结构合理；  2.报告/说明书表述清晰、规范等。 |  |
| 沟通表达 /团队协作  （满分30） | 项目包括：  1.独立学习能力；  2.与指导老师沟通；  3.团队协作，分工合作情况等。 |  | 成果（程序、流程图、电路图、实物等）  （满分30） | 项目包括  1.成果的合理与可行性；  2.成果的正确性和复杂度；  3.成果的特色与创新性。 |  |
| 答辩成绩  （满分20） | 项目包括：  1.对完成内容的描述；  2.对问题的回答情况。 |  |
| 平时成绩合计（百分制） | |  | 期末成绩合计（百分制） | |  |
| 课程平时成绩（五级制）： | |  |  | 课程期末成绩（五级制） | |  |

|  |
| --- |
| 评语：  指导教师签字：  2025年1月 10 日 |

**广 州 城 市 理 工 学 院**

**任 务 书**

1. 实践内容：

本实践课程聚焦自然语言处理核心技术，通过理论结合项目实战，掌握文本预处理、词向量建模、序列标注、情感分析、机器翻译等关键技术。具体实践内容见学习通说明文档。

2.任务要求：

（1）完成任务所需的算法设计。

（2）完成完成实现功能的程序设计并调试。

（3）答辩，完成课程报告的撰写，每人一份，按格式要求撰写，正文不少于10页。

3.实践教学基本要求：

（1）实践期间，学习态度要端正，主动学习各种技能。

（2）实践中要认真做好笔记，保存好数据和设计的成果。

（3）根据指导老师的要求，认真完成项目内容及撰写说明书和答辩。

（4）说明书要按书写规划，格式正确，表达清晰，层次分明，重点突出。

（5）牢固树立安全第一的思想，坚决消除安全隐患，做好防火安全、危险化学品安全、饮食卫生安全、操作安全等工作。实践期间，对违反安全规则和学校纪律，造成自身伤害者，由本人负责；对造成国家或他人财产损失或他人伤害等，应承担经济或法律责任；对违法违纪造成严重不良影响者，加重处分。

# 摘 要

*摘要标题为：三号、宋体、加粗*

*摘要正文为：小四号，宋体，行距为固定值20磅。段首行缩进2个汉字*

**什么是摘要？**

**摘要是论文内容不加注释和评论的简单陈述。**

**摘要应具有独立性和自含性，即不阅读论文的全文，便能获得必要的信息，摘要中有数据、有结论，是一篇完整的短文，可以独立使用，可以引用，可以用于推广。**

**摘要的内容应包括与论文同等量的主要信息，供读者确定有无必要阅读全文，也供文摘等二次文献引用。摘要的重点是成果和结论。**

**摘要中不用图、表、化学结构式、非公知公用的符号和术语。**

关键词：

**什么是关键词？**

**关键词是为了文献标引工作从论文中选取出来用于表示全文主题内容信息款目的单词或术语。一般3~5个。**

**关键词**：多变量系统；预测控制；环境试验设备

*小4号黑体加粗 关键词3～5个，宋体小4号字；关键词之间用分号隔开，最后一个关键词不打标点符号*

目 录

[摘 要 I](#_Toc184224381)

[第一章 STM32开发板和舵机介绍 1](#_Toc184224382)

[1.1 正文层次 1](#_Toc184224383)

[1.2 正文字号、字体、行距 1](#_Toc184224384)

[1.3 插图和插表 1](#_Toc184224385)

[第二章 相关原理介绍 2](#_Toc184224386)

[2.1 正文层次 2](#_Toc184224387)

[2.2 正文字号、字体、行距 2](#_Toc184224388)

[2.3 插图和插表 2](#_Toc184224389)

[第三章 实现过程 3](#_Toc184224390)

[3.1 正文层次 3](#_Toc184224391)

[3.2 正文字号、字体、行距 3](#_Toc184224392)

[3.3 插图和插表 3](#_Toc184224393)

[总结 4](#_Toc184224394)

# 中文分词

## 中文分词概述

中文的分词是指将连续的汉字序列按照合理的语序进行分割，是自然语言处理的基础环节。分词就是将连续的字序列按照一定的规范重新组合成词序列的过程。中文没有明确的单词边界标记，因此算法需要自动识别单词边界。该技术广泛应用于搜索引擎、机器翻译和情感分析等领域，其准确性直接影响下游任务的性能。

## 基于词表的分词算法

基于词表的分词算法通过预定义词典进行词语匹配，属于规则驱动的分词方法。其核心思想是通过最长匹配原则在词典中查找候选词，具有实现简单、时间复杂度低的优点，但依赖词典完备性，难以处理未登录词和新词。以下是三个常见的分词算法的介绍。

### 正向最大匹配法

正向最大匹配法（FMM）从左向右扫描文本，优先匹配词典中最长的候选词。算法维护动态起始指针，每次截取max\_len长度子串进行词典匹配，若匹配失败则逐步缩短子串长度直至找到合法词语或单字词。该算法符合人类从左到右的阅读习惯，但对右边界歧义

处理能力较弱。

### 逆向最大匹配法

逆向最大匹配法（RMM）从右向左处理文本，优先匹配句尾方向的最长候选词。通过倒序扫描和结果前插操作维护词语顺序，其核心机制与FMM对称。实验表明，RMM在处理某些复合词（如"野生动物园"）时更具优势，但对左边界歧义处理存在局限性。

### 双向最大匹配法

双向最大匹配法通过融合FMM和RMM结果提升准确率。采用"词数最少优先"和"单字词最少优先"的双重决策机制：当两种方法分词数量相同时选择单字词较少的结果，数量不同时选择词语总数较少的结果。该方法在保持规则方法效率优势的同时，显著降低了单一方向匹配的错误传播风险。

## Jieba分词

jieba分词是国内使用人数最多的中文分词工具。jieba的主要功能是做中文分词，可以进行简单分词、并行分词、命令行分词，当然它的功能不限于此，目前还支持关键词提取、词性标注、词位置查询等。不仅如此，Jieba分词依靠中文词库，利用一个中文词库，确定汉字之间的关联概率，对汉字间概率大的组成词组，形成分词结果，除了分词，用户还可以添加自定义的词组。是一个很好用的中文分词工具。

### 基于N-gram模型的中文分词

1. Gram是大词汇连续语音识别中常用的一种语言模型，对中文而言，我们称之为汉语语言模型(CLM, Chinese Language Model)。汉语语言模型利用上下文中相邻词间的搭配信息，可以实现到汉字的自动转换，
2. gram模型基于马尔可夫假设，认为第n个词的出现仅与前n-1个词相关。在分词中，二元语法(Bigram)模型通过计算词序列的联合概率实现分词优化，其概率公式为：

应用时，通过统计计算语料库中邻接词的共发生频率来计算转移概率。例如，对于“人工智能发展”这句话，模型计算“人工/智能/发展”和“人工智能/发展”这两个切入点的概率，并选择更高的切入点。该方法可以有效地处理简单的歧义，但对于长距离依赖处理具有限制。

### 基于隐马尔可夫模型的中文分词

隐马尔可夫模型（Hidden Markov Model, HMM）是一种基于概率的序列建模方法，广泛应用于中文分词任务。其核心思想是通过隐含状态序列和观测序列之间的概率关系，推断出最合理的分词结果。以下从模型定义、状态设计、训练过程和解码方法四个方面详细说明其实现原理。

隐马尔可夫模型(HMM)将分词建模为五元组：

λ=(S,V,π,A,B)

这五组分别是

状态集合S：{B(词首), M(词中), E(词尾), S(单字词)}

观测序列V：中文字符集合

初始概率π：各状态初始分布

状态转移矩阵A：B→M→E等转移概率

发射矩阵B：状态生成字符的概率

HMM分词的实现通过是Viterbi算法求解最优状态序列：给定字符序列O=(o₁,o₂,...,oₙ)，计算使P(Q|O)最大的状态序列Q=(q₁,q₂,...,qₙ)。

例如对"广州城市理工学院"，模型通过状态转移路径识别"广州/城市/理工学院"或"广州市/城市理工/学医"等不同切分。对应的HMM状态序列为：

广（B）州（E）城（B）市（E）理（B）工（M）学（M）院（E）

广州由B-E构成一个二字词；城市由B-E构成一个二字词；理工学院由B-M-M-E构成一个四字词；

### Jieba分词原理与流程

Jieba分词是中文自然语言处理领域广泛使用的高效分词工具，其核心思想是 “词典匹配为主，统计模型为辅” 的混合架构。通过结合规则方法与统计方法，Jieba在保证效率的同时提升了未登录词，如人名、地名、新词的识别能力。Jieba采用"词典+HMM"的混合架构，有前缀词典、有向无环图DAG、动态规划和我们之前提到的隐马尔可夫模型即HMM。

在前缀词典里面，Jieba分词采用了哈希表，而非传统Trie树，在存储词语及其词频时，实现了O(1)时间复杂度的快速查询；

有向无环图DAG中，主要是对输入句子中的每个字符节点，向后查找所有可能的词语组合。例如“人工智能”的有向无环图结构就是：

0: [0,1,2,3] # "人"、"人工"、"人工智"、"人工智能"

1: [1] # "工"

2: [2] # "智"

3: [3] # "能

动态规划也就是最大路径概率，其权重计算是基于词频计算路径累计权重，公式为：

最大路径概率公式录找到从句子起点到终点的最大权重路径，让分词速度加快。

隐马尔可夫模型，也叫HMM，可以使用{B, M, E, S}标注未登录词的分词边界，用于处理词典未覆盖的连续单字，可以自己推断出最合理的分词结果

Jieba分词流程分为5步，分别是初始化与词典加载、构建有向无环图、动态规划计算最优路径、切分未登录词和结果融合与输出

初始化与词典加载就是加载默认词典及用户自定义词典，然后统计总词频并归一化用于动态规划计算。

构建有向无环图，对输入句子生成所有可能的词语切分路径。

动态规划计算最优路径，从右向左反向计算每个节点的最大权重。

切分未登录词即HMM补全，在动态规划路径中识别到存在连续未在词典中出现的单字后触发HMM，根据状态序列{B, E}切分为/B和/E。

最后结果融合与输出，合并词典切分结果和HMM切分结果。

# XX加工/制作/调试（题目仅供参考）

## 正文层次

正文所有章节按“第一章、第二章、第三章……（换章时必须换页）；1.1、1.2、1.3……；1.2.1、1.2.2、1.2.3……”编排。各层次题序及标题不得置于页面的最后一行（孤行）。

## 正文字号、字体、行距

第一级标题用三号，宋体，加粗，左右居中，上下空一行；

第二级标题用小三号，宋体，加粗，靠左，上下空一行；

第三级标题用四号，宋体，加粗，靠左，不空行；

正文用小四号，宋体，行距为固定值20磅。段首行缩进2个汉字。

## 插图和插表

插图（表）的标题和图（表）内文字：五号，宋体。

插图（表）必须有图（表）题（由号和名组成）。图（表）号按章排序，如第一章第一图的号为“图1-1”，第二章第一幅图的号为“图2-1”。图题置于图下方，表题置于表上方。

插图与其图题不得拆开排写于两页。图中若有分图时，分图号用a)、b)等置于分图之下。有数字标注的坐标图，必须注明坐标单位。

表格不加左、右列线；表内数字空缺的格内加“——”字线；如某个表格需要换页接排，在随后的各页上应重复表的编排，编号后跟标题和（续）。如：表2-1 加入激素后的实验结果比较（续）。

# 整车联调测试（题目仅供参考）

## 正文层次

正文所有章节按“第一章、第二章、第三章……（换章时必须换页）；1.1、1.2、1.3……；1.2.1、1.2.2、1.2.3……”编排。各层次题序及标题不得置于页面的最后一行（孤行）。

## 正文字号、字体、行距

第一级标题用三号，宋体，加粗，左右居中，上下空一行；

第二级标题用小三号，宋体，加粗，靠左，上下空一行；

第三级标题用四号，宋体，加粗，靠左，不空行；

正文用小四号，宋体，行距为固定值20磅。段首行缩进2个汉字。

## 插图和插表

插图（表）的标题和图（表）内文字：五号，宋体。

插图（表）必须有图（表）题（由号和名组成）。图（表）号按章排序，如第一章第一图的号为“图1-1”，第二章第一幅图的号为“图2-1”。图题置于图下方，表题置于表上方。

插图与其图题不得拆开排写于两页。图中若有分图时，分图号用a)、b)等置于分图之下。有数字标注的坐标图，必须注明坐标单位。

表格不加左、右列线；表内数字空缺的格内加“——”字线；如某个表格需要换页接排，在随后的各页上应重复表的编排，编号后跟标题和（续）。如：表2-1 加入激素后的实验结果比较（续）。

# 总结

正文所有章节按“第一章、第二章、第三章……（换章时必须换页）；1.1、1.2、1.3……；1.2.1、1.2.2、1.2.3……”编排。各层次题序及标题不得置于页面的最后一行（孤行）。

第一级标题用三号，宋体，加粗，左右居中，上下空一行；

第二级标题用小三号，宋体，加粗，靠左，上下空一行；

第三级标题用四号，宋体，加粗，靠左，不空行；

正文用小四号，宋体，行距为固定值20磅。段首行缩进2个汉字。

插图（表）的标题和图（表）内文字：五号，宋体。

插图（表）必须有图（表）题（由号和名组成）。图（表）号按章排序，如第一章第一图的号为“图1-1”，第二章第一幅图的号为“图2-1”。图题置于图下方，表题置于表上方。

插图与其图题不得拆开排写于两页。图中若有分图时，分图号用a)、b)等置于分图之下。有数字标注的坐标图，必须注明坐标单位。

表格不加左、右列线；表内数字空缺的格内加“——”字线；如某个表格需要换页接排，在随后的各页上应重复表的编排，编号后跟标题和（续）。如：表2-1 加入激素后的实验结果比较（续）。