《自然语言处理原理与应用（Principle and Application of Natural Language Processing)》教学大纲

制定时间：2024 年 2 月

一、课程基本信息

（一）适用专业：智能科学与技术

（二）课程代码：3ZN1036A

（三）学分/课内学时：3 学分/48 学时

（四）课程类别：专业教育

（五）课程性质：选修/理论课

（六）先修课程：机器学习、深度学习、概率论与数理统计

（七）后续课程： 机器翻译，语音识别技术，推荐系统

二、课程教学目标

本课程将学习人工智能的一大具体应用方向：自然语言处理。课程不仅会讲解自然语言处理的主要任务和如何基于不同研究范式实现相关任务（包括知识库、统计、神经网络等），而且会重点以案例的形式讲解如何基于实际的自然语言处理框架，针对不同的应用场景进行解决应用问题。使学生能快速具备自然语言处理问题求解的基本思想和初步的自然语言处理软件开发能力。

（一）具体目标

目标 1：了解自然语言处理的主要任务，包括分词，序列标注，句法分析，文本分类，文本生成，命名实体识别，信息检索，机器翻译等。

目标 2：掌握基于规则或机器学习/深度学习的相关技术； 具备针对不同的自然语言处理应用场景选择对应的技术的能力；具备自然语言处理问题求解的基本思想和初步的自然语言处理软件开发能力。

（二）课程目标与毕业要求的对应关系

毕业要求	毕业要求指标点	课程目标	教学单元	评价方式
1. 能够应用数学、自然科学和工程科学的基本原理，识别、表达、并通过文献研究分析智能系统中的复杂工程问题，以获得有效结论。	观测点2.5：能运用基本原理分析实际工程的影响因素，证实解决方案的合理性。	目标 1	自然语言处理绪论, NLP 相关工具及数学基础, 分词, 序列标注, NLP 应用任务	课内实验 期末考试

2. 具有自主学习和终身学习的意识，有不断学习和适应发展的能力。	观测点12.1：具有自主学习和终身学习的意识，具备终身学习的知识基础和自主学习的方法；	目标 2	序列标注，NLP应用任务，期末考核	课内实验 期末考试

三、教学内容与方法

（一）教学内容及要求

序号	教学单元	教学内容	学习产出要求	推荐学时	推荐教学方式	支撑课程目标	备注
1	自然语言处理绪论	自然语言处理发展历史； 具体任务； 应用案例 课程目标；	了解自然语言处理的历史与前景展望，了解相关任务和应用案例； 了解本课程的学习目标 教学单元安排	3	讲授 案例	目标 1	
2	NLP 相关工具及数学基础	Hanlp； Nltk；jibea	搭建开发环境； 掌握必要的编程基础操作； 掌握必要的数学相关基础；	3	讲授 案例 实验	目标 1	
3	分词	**知识要点：** 词典 匹配法分词 n-gram	了解基于规则的分词； 了解字典树； 了解基于统计的分词； 了解 n 元语法； 掌握分词工具 jieba； 熟悉分词评价指标	8	讲授 案例 实验	目标 1	
4	序列标注	**知识要点：** 隐马尔科夫 HMM； 结构化感知机； 条件随机场 CRF	实现基于 HMM 的序列标注； 理解维特比算法； 实现基于结构化感知机的序列标注； 实现基于 CRF	14	讲授 案例 实验	目标 1 目标 2	

序号	教学单元	教学内容	学习产出要求	推荐学时	推荐教学方式	支撑课程目标	备注
			的序列标注；				
5	NLP 应用任务	知识要点： 词性标注 POSTAG； 命名实体识别 NER； 信息抽取 IE； 文本聚类； 文本分类； 依存句法分析	基于机器学习模型实现词性标注； 基于机器学习模型实现 NER； 基于无监督学习实现 IE； 基于多种模型并灵活运用特征工程实现文本聚类/分类； 体会句法分析	18		目标 1 目标 2	
6	期末考核	课程总结 答疑 考核内容的说明　布置期末大作业	熟悉经典的自然语言处理任务； 能够针对应用任务设计相应的解决方案	2	讲授 案例	目标 2	

（二）教学方法

本课程注重理论教学与实验的结合，注重学生实践能力的培养，加强实验上机来巩固学生对自然语言处理相关技术的理解，通过实验锻炼学生动手解决自然语言处理相关问题的能力，实验还将以目前比较常见的应用案例为实例，使学生体会自然语言处理的具体作用，通过本课程的学习，学生将全面了解实现自然语言处理的技术，能够在实际项目的研究中运用这些技术加速工作，跟踪前沿的自然语言处理应用场景等，能够为学生从事自然语言处理乃至人工智能相关实践项目打下坚实的基础。

1.课堂讲授

（1）采用启发式教学，激发学生主动学习的兴趣，培养学生独立思考、分析问题和解决问题的能力，引导学生主动通过实践和自学获得自己想学到的知识。

（2）在教学内容上，系统讲授自然语言处理的基本理论、基本知识和基本方法，使学生能够系统掌握用于解决智能科学类专业工程复杂问题的专业基础知

识。

（3）在教学过程中采用电子教案，Jupyter Notebook 课件，多媒体教学与传统板书、教具教学相结合，提高课堂教学信息量，增强教学的直观性。

（4）理论教学与工程实践相结合，引导学生应用数学、自然科学和工程科学的基本原理，采用现代设计方法和手段，进行问题分析、综合与仿真，培养其识别、表达和解决智能类专业相关工程问题的思维方法和实践能力。

（5）课内讨论和课外答疑相结合，每周至少一次进行答疑。

2. 实验教学

实验教学是本课程中重要的实践环节，目的是培养学生运用实验方法研究解决智能类专业复杂工程问题的能力。课程必做实验 8 个，各实验要求学生独立或分组完成，并提交实验报告至少 4 次。

3. 专题研究

围绕本课程教学重点内容，设置专题研究环节，培养学生逐步具有应用机器学习，深度学习等技术解决自然语言处理的应用问题的能力，结合所研究课题进行报告和设计报告的撰写，并清晰陈述观点和回答问题的能力。

组织形式及要求如下：

（1）学生从教师给定的题目中选择或自主选题，以小组为单位进行，每个人的分工与责任需明确，并在报告中提供小组研讨情况记录及说明；

（2）选题应结合具体任务的需求，设计自然语言处理应用程序，给出设计成果，撰写研究报告，并进行陈述与答辩。

四、考核及成绩评定

（一）考核内容及成绩构成

课程目标	考核内容	成绩评定方式	成绩占总评分比例	目标成绩占当次考核比例	学生当次考核平均得分	目标达成情况计算公式
目标 1：了解自然语言处理的主要任务，包括分词，序列标注，句法分析，文本分类，文本生成，命名实体识别，信息检索，机器翻译等。	分词,序列标注等任务	实验	15%	100%	A_1	$\frac{\frac{A1}{100\%}\times 15\%+\frac{A2}{100\%}\times 35\%}{50}$
	阐述自然语言处理任务的实现原理	期末	35%	100%	A_2	
目标 2：掌握基于规则或机器学习/深度学习的相关技术；具备针对不同的自然语言处理应用场景选择对应的技术的能力；具备自然语言处理问题求解的基本思想和初步的自然语言处理软件开发能力。	分析自然语言处理实现的规则和机器学习技术并提交实验报告	实验	15%	100%	B_1	$\frac{\frac{B1}{100\%}\times 15\%+\frac{B2}{100\%}\times 35\%}{50}$
	实现一个自然语言处理的综合任务并提交报告	期末	35%	100%	B_2	
总评成绩（100%）= 实验（40%）+期末（60%）			100%	——	——	$\frac{\text{学生总评平均分}}{100}$

（二）实验考核成绩评定

1.支撑目标 1、目标 2，共占总评分 30%，目标 1 占 15%、目标 2 占 15%。对应目标的评分标准如下：

对应目标	**目标 1**：了解自然语言处理的主要任务，包括分词，序列标注，句法分析，文本分类，文本生成，命名实体识别，信息检索，机器翻译等。	**目标 2**：掌握基于规则或机器学习/深度学习的相关技术； 具备针对不同的自然语言处理应用场景选择对应的技术的能力；具备自然语言处理问题求解的基本思想和初步的自然语言处理软件开发能力。

考查点		实验内容	实验报告
占总成绩比例		15%	15%
评分标准	100%至90%	实验记录全部完成无遗漏，内容丰富、图文并茂，流程图数量足够且正确，实验方案有自己独到的思路与见解。实验记录全部完成无遗漏，内容丰富、图文并茂，流程图数量足够且正确，实验方案有自己独到的思路与见解。	有很强的总结实验和撰写报告的能力，实验报告内容完整、正确，有很好的分析与见解。文本表述清晰，书写工整，格式规范。
	89.9%至80%	实验记录比较完整，内容比较丰富、图文并茂，流程图数量足够且基本正确，实验方案有自己的思路与见解。	有较强的总结实验和撰写报告的能力，实验报告内容完整、正确，有较好的分析与见解。文本表述较为清晰，书写比较工整，格式规范。
	79.9至70%	实验记录比较完整，内容比较丰富，流程图数量足够且基本正确。	有良好的总结实验和撰写报告的能力，实验报告内容较完整、正确，有自己的分析与见解。文本表述较为清晰，书写较为工整，格式较为规范。
	69.9%至60%	实验记录基本完整，内容基本够，流程图数量基本够但有少量错误。	有一定的总结实验和撰写报告的能力，实验报告内容基本完整、正确，没有分析或见解。文本表述基本清晰，书写基本工整，格式基本规范。
	59.9%至0	实验记录未完成，内容不够，流程图数量不够、错误多。	总结实验和撰写报告的能力差，实验报告内容不完整、错误多。文本表述不清晰，书写潦草、格式不规范。

五、参考学习资料

（一）推荐教材：

1. 何晗 著 自然语言处理入门 ISBN:9787115519764 人民邮电出版社，2019

2. 斋藤康毅(日) 著 陆宇杰 译. 深度学习进阶-自然语言处理, ISBN：9787115547644. 北京：人民邮电出版社，2020.

（二）在线资源:

1. 《自然语言处理入门》 配套资源
   https://od.hankcs.com/
2 《深度学习进阶 自然语言处理》 配套资源 官方发布
   https://github.com/oreilly-japan/deep-learning-from-scratch-2

制订人： 罗雯涛　　审核人： 杨怡康