

课程大纲

COURSE SYLLABUS

1. 课程代码/名称 Course Code/Title	高级自然语言处理 / Advanced Natural Language Processing STA5007
2. 课程性质 Compulsory/Elective	专业选修课 Major Elective Courses
3. 开课单位 Offering Dept.	统计与数据科学系 Department of Statistics and Data Science
4. 课程学分/学时 Course Credit/Hours	48
5. 授课语言 Teaching Language	中英文 Chinese and English
6. 授课教师 Instructor(s)	陈冠华 Chen Guanhua
7. 开课学期 Semester	秋季 Fall
8. 是否面向本科生开放 Open to undergraduates or not	是
9. 先修要求 Pre-requisites	(如面向本科生开放, 请注明区分内容。 If the course is open to undergraduates, please indicate the difference.) 本科生选课要求先修 CS203/CS203B (数据结构与算法分析 Data Structures and Algorithm Analysis)

10. 教学目标 Course Objectives

(如面向本科生开放, 请注明区分内容。 If the course is open to undergraduates, please indicate the difference.)

通过课程的学习, 学生将掌握自然语言处理领域的基础知识, 学习解决相关问题的常用框架与方法。课程教学的内容包括语言模型, 表示学习, 大语言模型, 智能体, 多模态大模型等等。通过学习课程学生学习使用所学知识解决自然语言处理领域的典型实际问题。

This course will introduce the basics of NLP, covering standard frameworks for dealing with natural language as well as methods to solve different NLP problems. The covered topics include language modeling, representation learning, large language models, LLM-based agents, multimodal LLMs, and others. Students are expected to learn to solve typical natural language processing tasks at the end of this course.

11. 教学方法 Teaching Methods

通过本课程的学习, 学生预期可达到:

1. 理解自然语言处理的基础知识和相关任务。
2. 选择合适的模型和算法解决典型的自然语言处理任务。

After taking this course, the students should be able to:

1. Understand the basic concepts and tasks in natural language processing.

2. Solve typical natural language processing tasks with corresponding models and algorithms.

12. 教学内容

Course Contents

(如面向本科生开放, 请注明区分内容。 If the course is open to undergraduates, please indicate the difference.)

Section 1	<p>1. 自然语言处理简介 (3 学时) 自然语言处理的历史及发展, 并介绍典型任务。</p> <p>1 Introduction to natural language processing (3 hours) The history of natural language processing, introduction to typical tasks in NLP</p>
Section 2	<p>2. 神经网络的基础知识 (3 学时) 理解不同类型的神经网络及其优化算法。</p> <p>2 Neural networks basics (3 hours) Understanding the basics of different neural networks and their optimization methods.</p>
Section 3	<p>3. 循环神经网络和语言模型 (3 学时) 介绍循环神经网络及其在语言模型中的应用</p> <p>3 Recurrent neural networks and language models (3 hours) Introduction to recurrent neural networks and its applications in language models.</p>
Section 4	<p>4. 注意力机制与 Transformer 模型 (6 学时) 理解注意力机制, Transformer 模型和文本生成任务。</p> <p>4 Sequence labeling (6 hours) Understanding attention mechanism, Transformer model and text generation task.</p>
Section 5	<p>5. 预训练语言模型 (6 学时) 预训练语言模型和预训练并微调范式的核心思想</p> <p>5 Pretrained language models (6 hours) The core idea of pretrained language models and the pretraining-and-finetuning paradigm.</p>
Section 6	<p>6. 大语言模型 (15 学时) 从数据、训练方法和解码角度介绍大语言模型及其核心思想</p> <p>6 Large language models (15 hours) The core idea of large language models in terms of data, training method and inference.</p>
Section 7	<p>7. 基于大模型的多智能体 (6 学时) 介绍基于大模型的多智能体框架, 模型的自我进化和多智能体协作。</p> <p>7 LLM-based agents (6 hours) Understanding the basic technology of LLM-based agents, self-evolvment and collaboration among multiple agents.</p>
Section 8	<p>8. 多模态大模型 (3 学时) 理解多模态大模型的基础任务、设计理念和应用场景。</p> <p>8 Multimodal LLM (3 hours) Understanding the basic tasks, designs and applications of multimodal large language models.</p>
Section 9	<p>9. 自然语言处理的发展趋势 (3 学时) 自然语言处理的未来发展趋势</p> <p>9 Future trends in NLP (3 hours)</p>

Introduction to the future trends of NLP.

13. 课程考核

Course Assessment

(考核形式 Form of examination; 分数构成 grading policy; 如面向本科生开放, 请注明区分内容。 If the course is open to undergraduates, please indicate the difference.)

出勤 Attendance 5% + 随堂测验 quiz 5% + 作业 Homework 40% + 项目 Project 25% + 期末考试 Final Exam 25% (考试)

14. 教材及其它参考资料

Textbook and Supplementary Readings

Textbook:

Speech and Language Processing (3rd ed. draft), Dan Jurafsky and James H. Martin, 2021.

Supplementary Readings:

《自然语言处理：基于大语言模型的方法》，车万翔，郭江，崔一鸣，电子工业出版社，2025.

《大语言模型》，赵鑫，李军毅等，高等教育出版社，2024