

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, 文本分类, 序列标注, 机器翻译, 问答系统等等。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, text classification, sequence tagging, machine translation, question answering, 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.	<p>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.)</p> <table border="1"> <tr> <td data-bbox="194 322 552 490">Section 1</td> <td data-bbox="552 322 1455 490"> 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。 1 Introduction to natural language processing (3 hours) The history of natural language processing, introduction to typical tasks in NLP </td> </tr> <tr> <td data-bbox="194 490 552 685">Section 2</td> <td data-bbox="552 490 1455 685"> 2. 神经网络的基础知识 (6 学时) 理解不同类型的神经网络及其优化算法。 2 Neural networks basics (6 hours) Understanding the basics of different neural networks and their optimization methods. </td> </tr> <tr> <td data-bbox="194 685 552 880">Section 3</td> <td data-bbox="552 685 1455 880"> 3. 循环神经网络和语言模型 (6 学时) 介绍循环神经网络及其在语言模型中的应用 3 Recurrent neural networks and language models (6 hours) Introduction to recurrent neural networks and its applications in language models. </td> </tr> <tr> <td data-bbox="194 880 552 1111">Section 4</td> <td data-bbox="552 880 1455 1111"> 4. 序列标注 (3 学时) 学习常见的序列标注任务, 如命名实体识别和依存句法分析等。 4 Sequence labeling (3 hours) Learning about sequence labeling tasks such as named entity recognition, dependency parsing </td> </tr> <tr> <td data-bbox="194 1111 552 1341">Section 5</td> <td data-bbox="552 1111 1455 1341"> 5. 机器翻译 (9 学时) 理解注意力机制, Transformer 模型和机器翻译任务。 5 Machine translation (9 hours) Understanding attention mechanism, Transformer model and machine translation task. </td> </tr> <tr> <td data-bbox="194 1341 552 1559">Section 6</td> <td data-bbox="552 1341 1455 1559"> 6. 预训练语言模型 (12 学时) 预训练语言模型和预训练并微调范式的核心思想 6 Pretrained language models (12 hours) The core idea of pretrained language models and the pretraining-and-finetuning paradigm. </td> </tr> <tr> <td data-bbox="194 1559 552 1753">Section 7</td> <td data-bbox="552 1559 1455 1753"> 7. 问答系统 (3 学时) 传统问答系统和开放域问答系统的常见技术 7 Question answering (3 hours) Understanding the basic technology of question answering and open-domain question answering. </td> </tr> <tr> <td data-bbox="194 1753 552 1948">Section 8</td> <td data-bbox="552 1753 1455 1948"> 8. 多语言自然语言处理 (3 学时) 理解迁移学习和在不同 NLP 任务中的跨语言迁移问题 8 Multilingual NLP (3 hours) Understanding transfer learning and cross-lingual transfer in different NLP tasks. </td> </tr> <tr> <td data-bbox="194 1948 552 2107">Section 9</td> <td data-bbox="552 1948 1455 2107"> 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 9 Future trends in NLP (3 hours) Introduction to the future trends of NLP. </td> </tr> </table>	Section 1	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。 1 Introduction to natural language processing (3 hours) The history of natural language processing, introduction to typical tasks in NLP	Section 2	2. 神经网络的基础知识 (6 学时) 理解不同类型的神经网络及其优化算法。 2 Neural networks basics (6 hours) Understanding the basics of different neural networks and their optimization methods.	Section 3	3. 循环神经网络和语言模型 (6 学时) 介绍循环神经网络及其在语言模型中的应用 3 Recurrent neural networks and language models (6 hours) Introduction to recurrent neural networks and its applications in language models.	Section 4	4. 序列标注 (3 学时) 学习常见的序列标注任务, 如命名实体识别和依存句法分析等。 4 Sequence labeling (3 hours) Learning about sequence labeling tasks such as named entity recognition, dependency parsing	Section 5	5. 机器翻译 (9 学时) 理解注意力机制, Transformer 模型和机器翻译任务。 5 Machine translation (9 hours) Understanding attention mechanism, Transformer model and machine translation task.	Section 6	6. 预训练语言模型 (12 学时) 预训练语言模型和预训练并微调范式的核心思想 6 Pretrained language models (12 hours) The core idea of pretrained language models and the pretraining-and-finetuning paradigm.	Section 7	7. 问答系统 (3 学时) 传统问答系统和开放域问答系统的常见技术 7 Question answering (3 hours) Understanding the basic technology of question answering and open-domain question answering.	Section 8	8. 多语言自然语言处理 (3 学时) 理解迁移学习和在不同 NLP 任务中的跨语言迁移问题 8 Multilingual NLP (3 hours) Understanding transfer learning and cross-lingual transfer in different NLP tasks.	Section 9	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 9 Future trends in NLP (3 hours) Introduction to the future trends of NLP.
Section 1	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。 1 Introduction to natural language processing (3 hours) The history of natural language processing, introduction to typical tasks in NLP																		
Section 2	2. 神经网络的基础知识 (6 学时) 理解不同类型的神经网络及其优化算法。 2 Neural networks basics (6 hours) Understanding the basics of different neural networks and their optimization methods.																		
Section 3	3. 循环神经网络和语言模型 (6 学时) 介绍循环神经网络及其在语言模型中的应用 3 Recurrent neural networks and language models (6 hours) Introduction to recurrent neural networks and its applications in language models.																		
Section 4	4. 序列标注 (3 学时) 学习常见的序列标注任务, 如命名实体识别和依存句法分析等。 4 Sequence labeling (3 hours) Learning about sequence labeling tasks such as named entity recognition, dependency parsing																		
Section 5	5. 机器翻译 (9 学时) 理解注意力机制, Transformer 模型和机器翻译任务。 5 Machine translation (9 hours) Understanding attention mechanism, Transformer model and machine translation task.																		
Section 6	6. 预训练语言模型 (12 学时) 预训练语言模型和预训练并微调范式的核心思想 6 Pretrained language models (12 hours) The core idea of pretrained language models and the pretraining-and-finetuning paradigm.																		
Section 7	7. 问答系统 (3 学时) 传统问答系统和开放域问答系统的常见技术 7 Question answering (3 hours) Understanding the basic technology of question answering and open-domain question answering.																		
Section 8	8. 多语言自然语言处理 (3 学时) 理解迁移学习和在不同 NLP 任务中的跨语言迁移问题 8 Multilingual NLP (3 hours) Understanding transfer learning and cross-lingual transfer in different NLP tasks.																		
Section 9	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 9 Future trends in NLP (3 hours) Introduction to the future trends of NLP.																		

13.	<p>课程考核 Course Assessment</p> <p>(① 考核形式 Form of examination; ② .分数构成 grading policy; ③ 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.)</p> <p>出勤 Attendance 10% + 作业 Homework 40% + 项目 Project 25%+期末报告 Final Presentation 25% (考查)</p>
14.	<p>教材及其它参考资料 Textbook and Supplementary Readings</p> <p>Textbook:</p> <p>Speech and Language Processing (3rd ed. draft), Dan Jurafsky and James H. Martin, 2021.</p> <p>Supplementary Readings:</p> <p>《自然语言处理：基于预训练模型的方法》，车万翔，郭江，崔一鸣，电子工业出版社，2021. 《基于深度学习的自然语言处理》，邓力，刘洋等，清华大学出版社，2020</p>