

复旦大学课程教学大纲

院系: 大数据学院

日期: 2023 年 02 月 21 日

课程代码	DATA130030.01		
课程名称	自然语言处理		
英文名称	Natural Language Processing （NLP）		
学 分 数	3.0	周学时	3
课程性质	<input type="checkbox"/> 通识教育专项 <input type="checkbox"/> 核心课程 <input type="checkbox"/> 通识教育选修 <input type="checkbox"/> 大类基础 <input type="checkbox"/> 专业必修 <input checked="" type="checkbox"/> 专业选修 <input type="checkbox"/> 其他		
教学目的	本课程通过讲授和课程项目实践方式，使得学生掌握自然语言处理（NLP）的基本工具和主要方法，使学生对 NLP 具有比较系统和完整的认识，并简单的了解现代 NLP 的技术，从而提高对自然语言数据处理的能力，为学生更加深入了解 NLP 并从事 NLP 研究工作打下基础。		
基本内容简介	本课程主要讲授 NLP 的基本方法以及原理，其中主要包括语言模型、观点提取、文本分类、问题回答、文本语义对比、机器翻译，同时也教授 NLTK 工具的使用、预训练技术等。本课程的自然语言主要是以文本形式存储，其中包括网页文本、Wikipedia、邮件文本等。这些文本数据包含了大量的有用信息，本课程将使用 Python 编程语言。		
考核和评价方式	本课程注重实践，课程评价方式如下： 1. 课程签到（5%） 2. 课堂表现（课堂测试、提问）（5%） 3. 平时作业（编程作业为主）（50%） 4. 实践项目（45%）		
基本要求： 本课程要求学生具备以下基础知识和能力： 1. 能够读懂英文材料； 2. 能够使用 Python 编写程序； 3. 掌握了基本的算法和数据结构知识； 4. 掌握了基本的统计以及概率论基础知识。			
使用教材： [1] Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. draft), 2021, Prentice Hall, https://web.stanford.edu/~jurafsky/slp3/ （主要参考教材） [2] Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999, https://nlp.stanford.edu/fsnlp/			
授课方式： 本课程以主讲教师讲授为主，兼以课堂上学生汇报和讨论课程相关的项目内容。			

教学团队成员(答疑时间表)

周宝健, 主讲 (<https://baojian.github.io/>), 周二, 10:00-11:30am

陈桢, 助教, 21210980104@m.fudan.edu.cn

助教以及助教工作内容: 协助主讲为学生答疑; 帮助主讲批改作业(编程作业为主); 协助学生, 解决课程实验项目上所遇到的技术性难题。

教学内容安排 (按 54 学时共计 17 周, 具体到每节课内容):

本课程主要讲授自然语言处理相关的基本任务, 内容包括文本表示、语言模型、文本分类、文本聚类、中文分词、词性标注、机器翻译、事件抽取等。在介绍不同任务时, 引入传统的基于机器学习和概率图模型的解决方案; 同时, 本课程为学生介绍神经网络和深度学习的基础知识, 并为相关的自然语言处理任务提供基于深度学习框架的解决方案。

课次	时间	具体内容
一	02/22/2023	自然语言处理概述
二	03/01/2023	文本表示、语言模型
三	03/08/2023	文本分类、机器学习与文本分类
四	03/15/2023	词向量表示、神经网络的语言模型
五	03/22/2023	文本聚类
六	03/29/2023	卷积神经网络、情感分析
七	04/05/2023	中文分词和词性标注
八	04/12/2023	隐马尔可夫模型和词性标注
九	04/19/2023	循环神经网络, 循环神经网络与词性标注
十	04/26/2023	机器翻译、贝叶斯网络模型
十一	05/03/2023	注意力机制、注意力机制与机器翻译
十二	05/10/2023	句法分析、递归神经网络与句法分析
十三	05/17/2023	事件抽取
十四	05/24/2023	文本摘要
十五	06/31/2023	社交媒体文本分析
十六	06/07/2023	对话自动生成系统
十七	06/14/2023	项目汇报
十八	06/21/2023	项目汇报

课内外讨论或练习、实践、体验等环节设计: 本课程以课内讲授、课外实践的方法为主, 通过学生完成一个实践作业掌握自然语言处理的基础原理和应用技术。