

《NLP 实训》

课程教学大纲

一、课程基本信息

课程类型	总学时为学时数	☐ 理论课（含上机、实验学时）			
	总学时为周数	☒ 实习 ☐ 课程设计 ☐ 毕业设计			
课程编码	7303501	总学时	1 周	学分	1
课程名称	NLP 实训				
课程英文名称	Natural Language Processing Training				
适用专业	数据科学与大数据技术				
先修课程	(7299611) 自然语言处理				
开课部门	信息学院人工智能系（大数据）				

二、课程性质与目标

本课程为数据科学与大数据技术专业必修课。本课程为学生掌握自然语言处理实际应用技巧奠定实践基础，目的是让学生熟悉自然语言处理的基本概念、语言模型、概率图模型等方面的知识，掌握利用 Python 自然语言处理工具包的编程实践，了解中文分词、输入法、机器翻译等实现过程，培养学生实现自然语言处理任务的能力。

课程目标 1：学生应掌握 NLP 技术中的理论运用方法，可处理一些具体问题。

课程目标 2：学生应能基于 Python NLP 工具箱及语料库，从兴趣出发，从细节入手，培养学生创新能力、实践能力。

课程思政目标：结合工程训练，培养学生敢于创新的行动能力和团队精神。

三、课程教学基本内容与要求

1. 教学内容一

掌握基于 Python NLTK 工具箱的使用，编程计算一段文本的熵以及不同文本之间的交叉熵。

2. 教学内容二

掌握语料库的查询及使用方法，利用语料库计算一段文本中每个字、特定分词出现的概率、N 元条件概率。

3. 教学内容三

掌握语法模型的使用方法，编程实现一个简单的自动分词系统，利用语料库

给出一段文本的分词结果及正确率。

4. 教学内容四

掌握语法模型的使用方法，编程实现一个简单的自动词性标注系统，利用语料库给出一段文本的词性标注结果及正确率。

5. 教学内容五

完成实训总结报告。

四、 实践性教学内容的安排与要求

实践教学内容安排。

第 1 天. 先进行基于 Python NLTK 工具箱的使用教学。学生编程计算一段文本的熵以及不同文本之间的交叉熵。

第 2 天. 先进行语料库的查询及使用方法教学。学生利用语料库计算一段文本中每个字、特定分词出现的概率、N 元条件概率。

第 3 天. 先进行语法模型 1 的教学。学生编程实现一个简单的自动分词系统，利用语料库给出一段文本的分词结果及正确率。

第 4 天. 先进行语法模型 2 的教学。学生编程实现一个简单的自动词性标注系统，利用语料库给出一段文本的词性标注结果及正确率。

第 5 天. 学生完成实训总结报告。

场地与设备：实训首日在机房教学，后续在校园内登录校园网，通过个人计算机浏览器访问分配的虚拟机完成实训环节。

五、 教学设计与教学组织

教学重点：

1. 贝叶斯网络在自然语言处理中面临的问题
2. 马尔可夫模型能避免贝叶斯网络问题的原因
3. 马尔科夫链的图网络表示方法及概率计算方法
4. 隐马尔可夫模型概念

教学难点：

1. 如何用隐马尔可夫等图网络模型表示自然语言模型
2. 如何进行概率计算并应用于拼音输入法中的拼音串转汉字串任务

讲授主要内容：

1. 概率图模型概述

介绍图的最基本表示单位。引导提问：如何用图表达一个导航路径选择的问题

题；引出概率图、路径和距离的概念。对照多种语言模型，进行比较，铺垫概率图模型在优化计算量方面的价值。

常见的图模型，不同图模型的相同点与不同点，引出 DBN 和 HMM 的差异问题。

通过简单的贝叶斯网络例子，介绍如何将一个联合概率问题表达成图的形式并计算。引出面临的问题；描述在 NLP 中，潜在的约束条件是什么，以及对 DBN 网络的简化思路，引出 HMM 的价值所在。

2. 在自然语言处理研究中常用的概率图模型

马尔可夫模型的性质及概率表达形式。利用状态图计算转移概率的方法。隐马尔科夫模型的概念及概率表达方法。

3. 小结

主体教学方式、方法：

课件、板书与自研实训系统相结合，线上线下混合教学。引导学生组成小组分工合作，将复杂的实训题目分解为子任务。鼓励学生以实训课题框架为基础，自主分析正确率受限的原因并尝试优化，实现勇于尝试、敢于创新的教学目标。

六、 教材与参考资料

1. 教材

《统计自然语言处理》（第 2 版），宗成庆，清华大学出版社，2013 年 8 月出版，ISBN 号 978-7-302-31911-5。

2. 参考资料

《Python 和 NLTK 自然语言处理》（第 1 版），尼天·哈登尼亚（Nitin Hardeniya）雅各布·帕金斯，人民邮电出版社，2019 年 4 月出版，ISBN 号 978-7-115-50334-3

七、 课程考核方式与成绩评定标准

实训成绩采用百分制。成绩评定按照：课程实训过程（占 50%），实训总结报告（占 50%）打分。

具体评定成绩的标准是：

- 1、实训过程遵守课程安排，态度端正，完成实训任务具有创新意识；
- 2、实训报告的内容完整、实现思路清晰、理论方法使用正确；实践过程逻辑完整，没有抄袭现象；
- 3、实训结果正确，分析结论可靠，同时考察学生上课积极学习态度，积极践行社会主义核心价值观和团结友爱进步精神。

八、 大纲制(修)订说明

无。

大纲执笔人：李洋

大纲审核人：王彦平

开课系主任：王彦平

开课学院教学副院长：宋威

制（修）订日期：2022 年 2 月