**《自然语言处理》课程教学大纲**

执笔人：张致远 修订日期：2022年11月

**一、课程基本信息**

1．课程编号：M301161B

2．课程平台：专业教育平台

3．课程模块：专业核心必修课程

4．课程性质：必修

5．学时/学分：48/3

6．先修课程：概率论与数理统计（B），线性代数，人工智能基础

7．适用专业：信息工程

8．教学单位名称：电信学院网络与安全教学团队

**二、课程教学目标**

**课程目标1：**理解自言语言处理的基本概念与基本方法；掌握自然语言处理中的词向量的构造和使用方法；掌握自然语言处理中深度学习的基本原理和使用方法；掌握基于RNN，CNN等方法实现文本标注，文本分类，实体关系抽取等任务的方法。

**课程目标2：**能够基于开源框架平台，解决自然语言处理中某一类实际应用问题，包括文本生成、图像语义生成、情感分析、机器翻译、问答等。通过研究和解决实践问题，提升为社会创造价值的能力。

**三、课程目标和毕业要求的对应关系**

|  |  |  |
| --- | --- | --- |
| **毕业要求** | **毕业要求指标点** | **课程目标** |
| 2. 工程知识：能够应用数学、自然科学、工程基础和专业知识的基本原理，将信息复杂工程问题抽象为数学、物理问题，选择适当的模型进行描述，对模型进行分析求解。 | 2.3将数学、自然科学、工程基础和专业知识用于工程问题的计算分析。 | 1，2 |
| 6. 使用现代工具：能够选择与使用适合的现代信息处理技术资源和设计工具，对于信息复杂工程问题进行预测和模拟，并理解所用工具和技术资源的局限性。 | 6.2能够开发、选择与使用恰当的技术、资源和现代工具，进行复杂工程问题的预测与模拟。 | 1，2 |
| 9. 职业规范：具有人文社会科学素养，具有社会主义核心价值观和社会责任感，能够在工程实践中理解并遵守工程职业道德和规范，履行责任。 | 9.1 具备人文社会科学素养，理解应担负的社会责任，愿意为社会服务。 | 2 |

**四、课程思政育人目标**

**目标1.** 介绍课程技术发展史，特别是有关方面的国家重大科技成果，让学生们树立正确理想信念，服务于国家重大战略。

**目标2.** 介绍课程中科技工作者的成就和贡献，培养学生严谨务实的治学精神和探索精神。

**目标3.** 引入研究型教学，配合思政案例，启发辩证思维，培养学生工匠精神，提高学生动手实践能力。

**五、课程教学内容和要求**

总学时 48学时，讲课40学时，实验8学时

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **知识单元（章节）** | **知识点** | **教学要求** | **推荐学时** | **教学方式** | **支撑课程目标** |
| 1 | 绪论 | 自然语言处理的概念 | 掌握 | 2 | 讲授 | 2.3, 9.1 |
| 自然语言处理的研究内容 | 了解 |
| 基本问题和主要困难 | 了解 |
| 自然语言处理的深度学习方法 | 掌握 |
| 自然语言处理的研究现状 | 理解 |
| 2 | 实验 | Python基础及TensorFlow框架介绍 | 掌握 | 2 | 讲授、上机操作 | 2.3,6.2 |
| 3 | 隐马尔科夫模型与条件随机场 | 马尔科夫模型 | 理解 | 2 | 讲授 | 2.3,6.2 |
| 隐马尔科夫模型 | 掌握 |
| 条件随机场 | 掌握 | 1 |
| 中文分词与词性标注 | 掌握 | 2 |
| 4 | 词的分布式表示 | 词向量概念 | 掌握 | 1 | 讲授 | 2.3,6.2 |
| 基于SVD的隐式语义分解 | 掌握 | 1 |
| word2vec与Glove算法 | 掌握 | 2 |
| 神经语言模型 | 掌握 | 2 |
| 5 | 实验 | word2vec与词向量表示 | 掌握 | 2 | 讲授、上机操作 | 2.3,6.2 |
| 6 | 自然语言处理与卷积神经网络 | CNN介绍 | 掌握 | 2 | 讲授 | 2.3,6.2 |
| CNN原理 | 掌握 |
| CNN与CRF比较 | 理解 | 1 |
| 文本分类、文本标注、关系抽取 | 理解 | 2 |
| 7 | 实验 | CNN文本分类 | 掌握 | 2 | 讲授、上机操作 | 2.3,6.2 |
| 8 | 自然语言处理与循环神经网络 | RNN网络的基本构造 | 掌握 | 2 | 讲授 | 2.3,6.2 |
| BPTT算法 | 理解 |
| 梯度消失和梯度爆炸 | 掌握 |
| LSTM与GRU | 掌握 | 1 |
| 文本标注、文本分类，机器翻译 | 理解 | 2 |
| 9 | 实验 | RNN与文本生成 | 掌握 | 2 | 讲授、上机操作 | 2.3,6.2，9.1 |
| 10 | 自然语言的树状结构 | 依存句法分析 | 了解 | 2 | 讲授 | 2.3,6.2 |
| 句法分析器 | 了解 | 2 |
| stanford nlp与哈工大LTP | 理解 | 1 |
| 11 | 实验 | 自然语言处理中的LSTM | 掌握 | 2 | 讲授、上机操作 | 2.3,6.2，9.1 |
| LSTM的改进方法 | 掌握 | 2 |
| 7 | Attention方法 | Attention方法的引入 | 理解 | 1 | 讲授 | 2.3,6.2 |
| memory network | 了解 | 2 |
| 8 | state of art模型 | BERT与Transformer | 掌握 | 2 | 讲授 | 2.3,6.2，9.1 |
| 情感分析，阅读理解与关系抽取 | 理解 | 2 |
| 自然语言处理的新方向 | 了解 | 1 |

**六、课程教学安排**

课程以课堂授课为主，结合同学自学和阅读教材、参考资料、相关论文，加深对教学内容的了解，在教学过程中应注重充分调动学生的主动性和积极性，并介绍相关的开源软件。各部分结束后应留出一定的思考题及实践问题作为学生阅读教材和参考资料的线索和提纲。引导学生使用深度学习框架实现基于深度学习框架的自然语言处理流程，训练从数据处理到模型搭建的一体化能力，提高教学效果，为同学进一步学习、深造打下坚实基础。

**（一）．课堂讲授**

在教学过程中，教师应以建立概念、理解和掌握基本理论为基础，指出每章的重点和难点部分。上课时，适当引入实际案例并提出问题，引导学生分析讨论，以调动学生的主动性。本课程的设计并不会从复杂的数学角度对自言语言处理涉及到的理论和方法从头至尾的给出数学证明和推导，而是从实际应用的角度出发，深入浅出讲解自然语言处理中所用到的Word2Vec、CNN、LSTM、attention等基础方法，并结合文本分类、机器翻译、情感分析、实体抽取等实际应用场景，通过课上实验结果展示让学生们对于课程上讲述的自言语言处理模型的构建过程及解决问题的核心思路等加深理解，提升学生的综合工程实践能力。

**（二）．指导自学**

鉴于学时数限制，同时为了培养锻炼学生自学能力，对部分课程内容提出自学要求，并指导自学。自学不仅包括概率论、深度学习中的基础数学知识，还包括框架和工具的自学。有助于使学生课下进行实验，让学生们对于课程上讲述的自言语言处理模型的构建过程及解决问题的核心思路等加深理解，提升学生的综合工程实践能力。

**（三）．专题研究讨论**

本课程设置专题研究讨论的目的是培养学生主动学习和举一反三的能力，可以结合课上的理论学习，结合课下自学，找到自己感兴趣的点去深入研究学习，例如：自然语言处理基础理论分析证明、情感分析中的隐式情感挖掘等。专题讨论可以以小组的方式进行组织，结合教学内容进展设置适当的专题，包括实验、理论分析、新技术追踪等，可以以书面报告的形式提交。

**（四）．课程思政**

《自然语言处理》课程理论性较强、概念抽象、数学知识点多。针对这些特点，将细化课程思政主题，挖掘知识点中的隐含思政元素，将课程思政案例（见下表）有机融入相关知识点和教学环节中，通过案例讲解、学生提问、课后调研、课堂展示等方式，进行知识传授和能力培养。

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 章节名称 | 知识点 | 所挖掘的思政元素 |
| 1 | 1.0 自然语言处理概论 | 自然语言处理应用无处不在 | 【**百度语音助手、科大讯飞、百度飞桨**】  我国在自然语言处理领域有比较突出的成果，但与国际先进水平还有差距，尤其是离真正的工业场景应用还有差距 |
| 2 | 2.0 隐马尔可夫模型和条件随机场 | 判别式模型和生成式模型 | 【**全面分析，开拓思路】**  生成式模型和判别式模型既有联系又有区别，应该联系的、全面的去分析问题 |
| 3 | 4.语言模型 | 由语言模型到神经语言模型到序列模型 | 【**创新意识、主动探索**】  从简单的语言模型表示到考虑词语相似性到上下文因素到考虑语义相似性，技术就是在解决问题中不断进步的。需要有这样发现问题的能力，创新精神去主动探索 |
| 4 | 5.6.7深度学习与自然语言处理相关内容 | 深度学习解决自然语言处理问题 | 【**认识差距，不断超越**】  我们再理论算法研究中处于较领先地，但硬件工具上我们还比较落后，尤其芯片产业。因此要认识差距，不断超越。 |

**七、课程的考核**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **课程成绩构成及比例** | **考核环节** | | **目标分值** | **考核/评价细则** | | **对应的课程目标** |
| 平时成绩  100分  占总评成绩的50% | 随堂测验 | 1（设计） | 10 | 主要考查各种基于深度学习的自然语言处理模型 | 随堂测验1、2的卷面成绩均为10分，分别计入平时成绩。 | 1 |
| 2（设计） | 10 | 主要考查各种自然语言处理任务 | 1 |
| 实验 | | 30 | 1. 主要考核学生对每章节知识点的复习、理解和掌握程度； 2. 考察学生的动手能力，解决实际问题能力 3. 每次实验6分，取各次成绩的加和值作为此环节的最终成绩。 | | 2 |
| 期末考试  100分  占总评成绩的50% | 设计模块 | | 70 | 1. 考查学生对自然语言处理技术的掌握程度及应用熟练度； 2. 卷面成绩100分，以卷面成绩的50%计入课程总评成绩。 3. 主要考核自言语言处理模型、基本任务等内容。考试题型为：选择题、判断题、简答题和设计题等。 | | 1 |
| 研究模块 | | 30 | 2 |

根据以上三项计算每名学生的总成绩，

**综合成绩=随堂测试成绩+实验成绩+期末考试成绩×0.5**

**八、本课程与其它课程的联系与分工**

先修课程：概率论与数理统计（B），线性代数，人工智能基础

后续课程：硕士阶段自然语言应用相关课程，如《知识图谱技术》

**九、建议教材及教学参考书**

建议教材：

1. 李梦全，TensorFlow与自然语言处理应用，清华大学出版社，2019.
2. 图删·加内格拉达著，马恩驰、陆健译，TensorFlow自然语言处理，机械工业出社，2019.

教学参考书：

1. 约阿夫·戈尔德贝格（Yoav Goldberg） 著，车万翔，郭江，张伟男，刘铭 译，基于深度学习的自然语言处理，机械工业出社，2018。

**十、大纲审核人**

**十一、学院审核程序说明**

由信息工程系制定，负责本科教学工作的系主任审核，经学院教学指导委员会审核批准。

**十二、学院审定日期**

**附：达成度评价方法：**

课程目标达成度评价包括课程分目标达成度评价和课程总目标达成度评价，具体计算方法如下：





课程目标评价内容及符号意义说明如附表1，字母*A*、*B*、*C*则分别表示随堂测验、专题研究、期末考试的实际平均得分，其中，*A*= *A*1+*A*2 *，C*= *C*1+*C*2；平时成绩和期末成绩分别占总评成绩的50%。

附表1 课程评价考核基本信息表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **课程目标评价内容** | **随堂测验** | | **实验** | **期末考试** | | **课程总评成绩** |
| 1 | 2 | 设计模块 | 研究模块 |
| 目标分值 | 10 | 10 | 30 | 70 | 30 | 100 |
| 学生平均得分 | *A*1 | *A*2 | *B* | *C*1 | *C*2 | A1+ A2+B+ C1+ C2 |

课程目标达成度评价值计算具体说明如附表2。

附表2课程达成度评价计算方法

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **课程目标** | **考核环节** | **目标分值** | **学生平均得分** | **达成度计算示例** |
| 课程目标1 | 随堂测验1 | 10 | *A*1 | 课程目标1达成度=（*A*1+ *A*2 *+ 0.5 C*1）/ 55 |
| 随堂测验2 | 10 | *A*2 |
| 期末考试（设计模块） | 70 | *C*1 |
| 课程目标2 | 实验 | 30 | *B* | 课程目标2达成度=（*B* + 0.5 *C*2）/ 45 |
| 期末考试（研究模块） | 30 | *C*2 |
| 课程总体目标 | 总评成绩 | 100 | (*A*+*B*)+0.5*C* | 课程总体目标 =【(*A*+*B*)+0.5*C*】/ 100 |