

## 第一部分 人工智能引论与思维方法

### 第一周 人工智能引论：概率模型|语言自动生成模型

- 1.1 形式语言与语法树：语法树(Syntax)与语言表示
- 1.2 语法树的构建解析：通过语法树生成句子实例
- 1.3 如何判断语句的合理：语言模型(Language Model)的构建
- 1.4 语言模型与语法树的结合来生成对话
- 1.5 作业：基于Syntax Tree 和 Language Model 实现西部世界对话智能系统

### 统

- 1.5.1) 数据驱动的编程
- 1.5.2) Python 实现句子生成与语法解析

### 第二周 人工智能引论：智能搜索策略|机器学习

- 2.1 智能搜索方法论
  - 2.1.1) 搜索问题与决策问题
  - 2.1.2) 智能搜索典型问题分析(传教士过河、八皇后等问题)
- 2.2 智能搜索的实现
  - 2.2.1) 深度搜索，广度搜索与最优搜索
- 2.3 机器学习问题的引入，搜索为什么不能解决很多问题？为什么需要机器学习？

- 2.3.1) 梯度下降与函数拟合
- 2.3.2) 从0到1，构建你的第一个机器学习模型
- 2.3 作业：北京市地铁自动换乘
  - 2.3.1) 数据获取爬虫的建立
  - 2.3.2) 编写智能搜索Agent

### 第三周 如何衡量模型的好坏？机器学习的评价与指标

- 5.1 机器学习的历史发展与原理
  - 5.1.1) 机器学习的背景与原理
  - 5.1.1) 机器学习的主要流派
  - 5.1.2) 机器学习的现状分析
- 5.2 过拟合与欠拟合
  - 5.2.1) Bias和 Variance
  - 5.2.2) 模型能力的分析
  - 5.2.3) 数据能力的分析
  - 5.2.4) 过拟合与欠拟合的原理与策略
- 5.3 训练集，测试集，准确度
  - 5.3.1) 数据对机器学习模型的影响
  - 5.3.2) 训练集、测试集、准确度的关系

## 第二部分 深度学习理论与自然语言处理

### 第4周 深度学习初步

#### 3.1 神经网络3

##### 3.1.1) Loss函数, Backpropagation

##### 3.1.2) 梯度下降

##### 3.1.3) softmax, cossentropy

##### 3.1.4) Optimizer 优化器

#### 3.2 神经网络的实践分析

##### 3.2.1) 模型的稳定性

##### 3.2.2) 模型的可解释性

##### 3.2.3) 模型的运行分析

#### 3.3 作业：手动从零实现一个神经网络模型

(本作业中，我们会带领大家使用纯Python和Numpy，自己从零完成一个神经网络, mini-flow, 然后使用该网络完成一个实际的问题，据此，你可以了解到深度学习的根本原理。)

##### 3.3.1) 实现神经元

##### 3.3.2) 实现拓扑排序

##### 3.3.3) 实现 Backpropagation

##### 3.3.4) 实现神经元权重自动调整

##### 3.3.5) 利用完成的神经网络模型进行房价预测

### 第 5周 word2vec与文本表征

#### 4.1 word embedding与词向量

##### 4.1.1) 词向量的原理

##### 4.1.2) 哈夫曼树与 Negative Samples

##### 4.1.3) GloVe, CoVe, ELMO 等高级词向量方式

##### 4.1.4) Python 利用神经网络训练词向量的实例

#### 4.2 句子向量 (sentence embedding)

##### 4.2.1) 句子向量的使用场景与背景

##### 4.2.2) 句子向量的构建与评价标准

##### 4.2.3) Python 构建句子向量的实例

#### 4.3 词向量的高级用法

##### 4.3.1)利用词向量找到隐藏重要词汇

##### 4.3.2)利用词向量找到新词汇

#### 4.4 作业：使用词向量自动整理同义词

#### 4.1 实训三： 文本自动摘要系统的构建讲解与导引

- 自动摘要的问题背景与挑战
- 中文文本摘要遇到的问题

- 使用 TextRank 进行文本自动摘要的实现
- 使用 Sentence Embedding 句子向量进行文本自动摘要的实现
- 完整的文本摘要系统所需要的技术能力分析

## 第六周 如何处理区域性问题？CNN卷积神经网络的原理及其应用

### 6.1 卷积神经网络与 Spatial Invariant

#### 6.1.1) 卷积神经网络的历史背景

#### 6.1.2) 卷积神经网络空间平移不变形(Spatial Invariant)的原理

#### 6.1.3) 卷积神经网络与 weights sharing

#### 6.1.4) 卷积神经网络的原理及Python 实现

### 6.2 Pooling, Dropout 与 Batch Normalization

#### 6.2.1) Pooling

#### 6.2.2) Dropout

#### 6.2.3) Batch Normalization

### 6.3 CNN 的可视化

### 6.4 CNN 模型分析的演化：

#### 6.4.1) LeNet

#### 6.4.2) AlexNet

#### 6.4.3) GoogLeNet

#### 6.4.4) VGG, ResNet

#### 6.4.5) DenseNet

### 6.5 作业：进行萝莉和正太的分类

#### 6.5.1) Python 深度学习环境的搭建

#### 6.5.2) Keras, Tensorflow 的介绍与使用方法

#### 6.5.2) 使用 Keras 搭建CNN 模型模型

#### 6.5.3) 模型的调试与优化

#### 6.5.4) 模型的发布

## 第七周 如何处理时序问题？RNN循环神经网络及其应用

### 7.1)序列模型

#### 7.1.1) 时间序列问题的分析

#### 7.1.2) 时间序列模型存在的问题挑战

### 7.2)RNN 循环模型

#### 7.2.1) RNN 的原理

#### 7.2.2) RNN 的相关问题

### 7.3)LSTM 与 GRU

#### 7.3.1) LSTM的原理与实现

#### 7.3.2) GRU 的原理

#### 7.3.3) Python 进行 RNN 模型的最佳实践

### 7.4)RNN 训练的高级问题

## **7.5)Transfer Learning 迁移学习**

7.5.1) 迁移学习的背景

7.5.2) 迁移学习的方法

7.5.3) Python 实现迁移学习的最佳实践

第八周 深度学习的前沿是？深度学习高级问题详解

### **8.1 Seq2Seq, Transform, BERT**

8.1.1) Seq2Seq 的原理

8.1.2) Seq2Seq 中的搜索方法

8.1.3) Attention注意力机制

8.1.4) Python Seq2Seq模型的最佳实践

8.1.4) Transform 机制及其应用

8.1.5) BERT 原理及其应用

### **8.2 强化学习**

8.2.1) 强化学习的原理

8.2.2) 强化学习的常用方法与实例

8.2.3) 强化学习面临的挑战

**8.3 自动对话机器人、文本自动摘要生成、文本自动阅读理解、自动驾驶等深度学习高级问题**

**项目作业二：细粒度客户评论自动分类**

- 数据预处理过程
- 数据分析与整理
- 模型的分析与搭建
- 模型的调整与分析
- 模型的部署与发布

## **第三部分 经典机器学习方法**

第九周：除了深度学习，还有什么模型，为什么依然需要？经典机器学习模型详解：

### **9.1 经典机器学习模型**

9.1.1) 深度学习的瓶颈

9.1.2) 逻辑回归

9.1.3) 贝叶斯分类器

9.1.4) KNN模型，

9.1.5) SVM及其原理

9.1.6) 决策树

9.1.7 Python) 机器学习模型的最近实践

## **9.2 Ensemble 机器学习方法**

### **9.2.1) Ensemble 机器学习的原理**

### **9.2.2) Random Forest 随机森林**

### **9.2.3) XGBOOST模型**

## **第十周：经典机器学习二：**

### **10.1 非监督/半监督学习与聚类模型：**

#### **10.1.1) K-Means算法与实例**

#### **10.1.2) 层次聚类与实例**

#### **10.1.3) 基于 embedding 的聚类机器实例**

### **10.2 机器学习常见实践问题分析**

#### **10.2.1) 天气预测**

#### **10.2.2) 文本分类**

#### **10.2.3) 图像分类**

#### **10.2.4) 机器阅读理解**

#### **10.2.5) 博弈问题**

### **10.3 作业：实现贝叶斯分类器，依据药物说明书进行药物适应症自动识别**

## **第四部分 自然语言处理的关键问题**

## **第十一周 动态规划与编辑距离：**

### **11.1)优化问题**

#### **11.1.1) 优化问题的背景**

#### **11.1.2) 优化问题的现状与常用方法**

### **11.2)动态规划**

#### **11.2.1) 动态规划的原理**

#### **11.2.2) 动态规划的典型实例**

#### **11.2.3) Python 实现动态规划的最佳实践**

#### **11.2.4)编辑距离的原理**

#### **11.2.5)编辑距离的 python 实现**

### **11.3 作业：中文拼写错误自动纠正**

#### **11.3.1 语言模型的构建**

#### **11.3.2 编辑距离的计算**

#### **11.3.3 自动纠错算法的实现**

#### **11.3.4 Python源代码完整分析**

## **第十二周 自然语言理解的关键问题**

### **12.2 关键词提取**

#### **12.2.1) 关键字提取的主要方法与挑战**

#### **12.2.2) 基于频率的TFIDF**

- 12.2.3) 基于图关系的 Text-Rank
- 12.2.4) 基于机器学习的方法
- 12.2.5) 基于词向量与图网络的方法
- 12.2.6) Python 关键词提取的最近实践
- 12.3 实体识别**
  - 12.3.1) 实体识别的原理与现状
  - 12.3.1) 实体识别的应用场景
  - 12.3.2) Python 实体识别的最佳实践
- 12.4 依存分析**
  - 12.4.1) 依存分析的原理与现状
  - 12.4.2) Python 依存分析的最佳实践

### 第十三周：搜索引擎与文档检索：

- 13.1 自动检索系统**
  - 13.1.1) 搜索引擎与文档搜索的背景
  - 13.1.2) 基于关键字的文本搜索
  - 13.1.3) 布尔代数搜索
- 13.2 PageRank**
  - 13.2.1) PageRank 原理
  - 13.2.2) PageRank 的其他应用场景
  - 13.2.3) Python 实现大规模搜索系统的关键能力与算法实例

### 项目三：PDF 重点信息智能标准

- 数据获取，数据转换，数据标准化
- 词向量的构建
- 依存分析与实体识别，重要信息识别
- 搜索系统
- 综合实现

### 第十四周：知识图谱与自然语言处理

- 14.1 图谱简介**
  - 14.1.1 起源，发展脉络，公开图谱，业内状况
  - 14.1.2 图谱涉及的理论、技术
- 14.2 图谱冷启动（初始化）**
  - 14.2.1 获取数据
  - 14.2.2 数据存储
  - 14.2.3 数据清洗
  - 14.2.4 实体分类
- 14.3 实体抽取(NRE)**
  - 14.3.1 序列标注——规范，数据集，工具

- 14.3.2 序列标注——HMM, MEMM, CRF
- 14.3.3 序列标注——LSTM, Transformer, Bert
- 属性抽取（非结构化）
- 14.4.1 基于模板的方法——触发词、依存句法
- 14.4.2 监督学习——深度学习、PCNN
- 14.4.3 半监督学习——远程监督、Bootstrapping
  - 实体链接-合并
  - 图谱应用
- 14.5.1 检索
- 14.5.2 对话
- 14.5.3 推荐

## 第十五周：面向服务的对话机器人的构建讲解与导引（实训四）

- 15.1 对话机器人的历史背景
- 15.2 使用语法树进行对话的实现
- 15.3 意图分析(Intention Recognition)与识别
- 15.4 短文本相似度(Short Text Similarity)匹配
- 15.5 文本快速检索
- 15.6 对话机器人的整体架构分析

## 第十六周：目前人工智能与局限性、前沿 NLP 问题的现状及发展情况

- 16.1 学习能力迁移问题，样本迁移问题
- 16.2 机器学习的可解释性
- 16.3 非结构数据的处理的计算
- 16.4 经典 AI 模型的计算复杂性
- 16.5 AI 产业化面临的问题：
  - 16.5.1) 数据标注与结构化数据
  - 16.5.1) 问题定义与认识
  - 16.5.2) 少量数据与机器学习的不可行性
- 16.6.基于人类背景知识的常识推理与认知问题