

第一部分 基础篇 经典人工智能模型方法

第一周 人工智能导论与语法树，自动机理论

1.1 形式语言与语法树

- 1.1.1) 形式语言与正则表达式
- 1.1.2) 语法树与语言表示
- 1.1.3) 使用语法树自动生产语言

1.2 自动机理论

- 1.2.1) 有限自动机基础
- 1.2.2) 语法树解析
- 1.2.3) 语法树解析的实际例子

1.3 作业：基于Syntax Tree实现西部世界对话智能系统

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

第二周 智能搜索策略

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
- 2.2.2) 搜索剪枝问题

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 动态规划与线性优化：

3.1) 优化问题

- 3.1.1) 优化问题的背景
- 3.1.2) 优化问题的现状与常用方法

3.2) 动态规划

- 3.2.1) 动态规划的原理
- 3.2.2) 动态规划的典型实例
- 3.2.3) Python 实现动态规划的最佳实践

3.3) 线性优化

- 3.3.1) 线性优化的原理
- 3.3.2) 线性优化经典作业
- 3.3.2) Python 线性优化最近实践

3.4) 作业：上海市外卖小哥送餐路线规划问题

- 3.4.1) 问题复杂度分析
- 3.4.2) 使用动态规划解决实例
- 3.4.3) 使用线性规划解决实例

第四周 自然语言理解初步

4.1 词向量

- 4.1.1) 文本表示初步
- 4.1.2) 降维与 embedding 的原理;
- 4.1.3) 词向量初步知识
- 4.1.4) Python 词向量使用的最佳实践

4.2 关键词提取

- 4.2.1) 关键字提取的主要方法与挑战
- 4.2.2) 基于频率的TFIDF
- 4.2.3) 基于图关系的 Text-Rank
- 4.2.4) 基于机器学习的方法
- 4.2.5) 基于词向量与图网络的方法
- 4.2.6) Python 关键词提取的最近实践

4.3 实体识别

- 4.3.1) 实体识别的原理与现状
- 4.3.1) 实体识别的应用场景
- 4.3.2) Python 实体识别的最佳实践

4.4 依存分析

- 4.4.1) 依存分析的原理与现状
- 4.4.2) Python 依存分析的最佳实践

第五周：搜索引擎与文档检索：

5.1 自动检索系统

- 5.1.1) 搜索引擎与文档搜索的背景
- 5.1.2) 基于关键字的文本搜索
- 5.1.3) 布尔代数搜索

5.2 PageRank

- 5.2.1) PageRank 原理
- 5.2.2) PageRank 的其他应用场景
- 5.2.3) Python 实现大规模搜索系统的关键能力与算法实例

项目实训一：新闻人物言论自动提取 或 PDF 重点信息智能标准系统

- 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

第二部分 机器学习与深度学习

第六周 统计概率模型：

6.1 语言模型

- 6.1.1) 语言模型的历史背景与意义
- 6.1.2) 语言模型的种类

6.2 统计概率模型

- 6.2.1) 条件概率
- 6.2.2) Python统计语言概率的实现
- 6.2.3) Good-Turing Estimation

6.3 编辑距离与文本相似度

- 6.3.1) 文本相似度的主要方法
- 6.3.2) 编辑距离的原理
- 6.3.3) 编辑距离的 python 实现

6.4 作业：中文拼写错误自动纠正

- 6.4.1 语言模型的构建
- 6.4.2 编辑距离的计算
- 6.4.3 自动纠错算法的实现
- 6.4.4 Python源代码完整分析

第七周 经典机器学习一

7.1 机器学习的历史发展与原理

- 7.1.1) 机器学习的背景与原理
- 7.1.1) 机器学习的主要流派
- 7.1.2) 机器学习的现状分析

7.2 过拟合与欠拟合

- 7.2.1) Bias和 Variance
- 7.2.2) 模型能力的分析
- 7.2.3) 数据能力的分析
- 7.2.4) 过拟合与欠拟合的原理与策略

7.3 训练集，测试集，准确度

- 7.3.1) 数据对机器学习模型的影响
- 7.3.2) 训练集、测试集、准确度的关系

第八周：经典机器学习二：

8.1 经典机器学习模型

- 8.1.1) 回归和分类
- 8.1.2) 逻辑回归
- 8.1.3) 贝叶斯分类器
- 8.1.4) KNN模型，
- 8.1.5) SVM
- 8.1.6) 决策树
- 8.1.7 Python) 机器学习模型的最近实践

8.2 Ensemble 机器学习方法

- 8.2.1) Ensemble 机器学习的原理
- 8.2.2) Random Forest 随机森林
- 8.2.3) XGBOOST模型

第九周：经典机器学习三：

9.1 非监督/半监督学习与聚类模型：

- 9.1.1) K-Means算法与实例
- 9.1.2) 层次聚类与实例
- 9.1.3) 基于 embedding 的聚类机器实例

9.2 机器学习常见实践问题分析

- 9.2.1) 天气预测
- 9.2.2) 文本分类
- 9.2.3) 图像分类
- 9.2.4) 机器阅读理解
- 9.2.5) 博弈问题

9.3 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

第十周 深度学习初步

10.1 神经网络

- 10.1.1) Loss函数, Backpropagation
- 10.1.2) 梯度下降
- 10.1.3) softmax, cossentropy
- 10.1.4) Optimizer 优化器

10.2 神经网络的实践分析

- 10.2.1) 模型的稳定性
- 10.2.2) 模型的可解释性
- 10.2.3) 模型的运行分析

10.5 作业：手动从零实现一个神经网络模型

- 10.3.1) 实现神经元
- 10.3.2) 实现拓扑排序
- 10.3.3) 实现 Backpropagation
- 10.3.4) 实现神经元权重自动调整
- 10.3.5) 利用完成的神经网络模型进行真实机器学习任务

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

第十一周 word2vec

11.1 word embedding与词向量

- 11.1.1) 词向量的原理
- 11.1.2) 哈夫曼树与 Negative Samples
- 11.1.3) GloVe, CoVe, ELMo 等高级词向量方式
- 11.1.4) Python 利用神经网络训练词向量的实例

11.2 句子向量

- 11.2.1) 句子向量的使用场景与背景
- 11.2.2) 句子向量的构建与评价标准
- 11.2.3) Python 构建句子向量的实例
- 11.3 词向量的高级用法**
- 11.3.1) 利用词向量找到隐藏重要词汇
- 11.3.2) 利用词向量找到新词汇
- 11.4 作业：使用词向量自动整理同义词**

第十二周 CNN卷积神经网络

- 12.1 卷积神经网络与 Spatial Invariant**
- 12.1.1) 卷积神经网络的历史背景
- 12.1.2) 卷积神经网络空间平移不变形(Spatial Invariant)的原理
- 12.1.3) 卷积神经网络与 weights sharing
- 12.1.4) 卷积神经网络的原理及Python 实现
- 12.2 Pooling, Dropout 与 Batch Normalization**
- 12.2.1) Pooling
- 12.2.2) Dropout
- 12.2.3) Batch Normalization
- 12.3 CNN 的可视化**
- 12.4 经典 CNN 模型分析：**
- 12.4.1) LeNet
- 12.4.2) AlexNet
- 12.4.3) GoogLeNet
- 12.4.4) VGG, ResNet
- 12.4.5) DenseNet
- 12.5 作业：进行萝莉和正太的分类**
- 12.5.1) Python 深度学习环境的搭建
- 12.5.2) Keras, Tensorflow 的介绍与使用方法
- 12.5.2) 使用 Keras 搭建CNN 模型模型
- 12.5.3) 模型的调试与优化
- 12.5.4) 模型的发布

第十三周 RNN循环神经网络

- 13.1) 序列模型**
- 13.1.1) 时间序列问题的分析
- 13.1.2) 时间序列模型存在的问题挑战
- 13.2) RNN 循环模型**
- 13.2.1) RNN 的原理
- 13.2.2) RNN 的相关问题
- 13.3) LSTM 与 GRU**
- 13.3.1) LSTM的原理与实现
- 13.3.2) GRU 的原理
- 13.3.3) Python 进行 RNN 模型的最佳实践

- 13.4) RNN 训练的高级问题
- 13.5) Transfer Learning 迁移学习
 - 13.5.1) 迁移学习的背景
 - 13.5.2) 迁移学习的方法
 - 13.5.3) Python 实现迁移学习的最佳实践

第十四周 深度学习高级问题

- 14.1 Seq2Seq, Transform, BERT
 - 14.1.1) Seq2Seq 的原理
 - 14.1.2) Seq2Seq 中的搜索方法
 - 14.1.3) Attention 注意力机制
 - 14.1.4) Python Seq2Seq 模型的最佳实践
 - 14.1.4) Transform 机制及其应用
 - 14.1.5) BERT 原理及其应用
- 14.2 强化学习
 - 14.2.1) 强化学习的原理
 - 14.2.2) 强化学习的常用方法与实例
 - 14.2.3) 强化学习面临的挑战
- 14.3 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

第十五周：面向服务的智能客户机器人与新闻自动摘要生成

- 15.1 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
- 15.2 实训四： 面向服务的对话机器人的构建讲解与导引
 - 对话机器人的历史背景
 - 使用语法树进行对话的实现
 - 意图分析与识别
 - 文本相似度匹配
 - 文本快速检索
 - 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