

# 课程方向：人工智能典型案例与实战

## 课程目标与主要内容：

1. 全面介绍人工智能的基本原理和方法；系统讨论知识图谱、智能搜索、不确定性推理等技术；重点分析常用的热点机器学习方法，卷积、循环、长短时记忆网络深度学习方法，以及深度强化学习方法。
2. 以案例为主线，系统掌握重要机器学习及深度学习方法的工具使用及实现技术
3. 从本质去理解人工智能建模所需要的思维
4. 学习人工智能在企业中的落地以及面临的挑战
5. 了解目前业界需要的人工智能人才的技能要求
6. 考评合格者，可获得人工智能典型案例与实战培训结业证书



王万森

教授，博士，享受国务院政府特殊津贴。原教育部高等学校计算机类专业教指委“智能科学与技术”专业教学指导工作组副组长；现任中国人工智能学会教育工作委员会主任；曾任中国人工智能学会秘书长。



李文哲

美国南加州大学博士，北京贪心科技创始人兼CEO，曾任凡普金科集团(爱钱进)首席数据科学家、美国亚马逊和高盛的高级工程师

## 课程大纲 (共计 48 学时, 6 天)

### 第 1 天 | 第一章: 人工智能概述, 确定性、不确定性知识系统, 空间搜索技术 (4 学时)

- 人工智能概述 (重要概念、应用领域、技术热点)
- 新一代人工智能基础理论及关键技术
- 确定性知识系统, 知识图谱
- 不确定知识系统 (可信度、主观 Bayes、证据理论、模糊推理等)
- 空间搜索技术: A 算法, A\* 算法,  $\alpha$ - $\beta$  剪枝

### 第二章: 演化搜索, 机器学习 (4 学时)

- 演化搜索: 遗传算法
- 线性回归
- 逻辑回归算法
- 决策树与 ID3 算法
- 支持向量机
- 集成学习与随机森林

### 第 2 天 | 第三章: 大数据分析, 神经网络与深度学习 (4 学时)

- 大数据分析概述: 关联、聚类、分类、降维等
- 神经网络与深度学习基础
- BP 网络学习
- 深度卷积网络学习

### 第四章: 深度学习, 深度强化学习, 人工智能现状分析 (4 学时)

- 循环网络学习
- 长短时记忆网络学习
- 深度强化学习
- 群体智能 (蚁群算法、粒群算法)
- 混合智能与人机协同交互概述
- 专家系统与智能决策概述
- 自然语言处理概述
- 高级机器学习方法概述
- 类脑智能概述

**第 3 天 | 第五节：广告点击率预测系统搭建（4 学时）**

- 广告点击率预测介绍
- 二分类问题以及算法
- 逻辑回归于梯度下降法
- 模型复杂度与过拟合现象
- L1, L2 正则介绍
- 交叉验证
- 特征工程
- 特征的处理
- AUC 评估标准

**代码与案例实践**

项目描述：基本上很多的互联网企业的主要收入来自于广告。其中对收入影响最大的就是广告的点击率，而且需要根据预测的点击率来优化广告列表的展示。

- 数据预处理
- 特征工程
- 选择合适的特征
- 逻辑回归模型的搭建
- 利用正则去避免过拟合现象
- 模型的评估

**项目学习目标**

- 掌握特征工程的技术
- 掌握机器学习建模的整套流程
- 深入理解逻辑回归模型以及设计原理
- 深入理解过拟合现象，以及如何在实际应用中避免此类现象

**第 3 天 | 第五节：广告点击率预测系统搭建（4 学时）**

- 广告点击率预测介绍
- 二分类问题以及算法
- 逻辑回归于梯度下降法
- 模型复杂度与过拟合现象
- L1, L2 正则介绍
- 交叉验证
- 特征工程
- 特征的处理
- AUC 评估标准

**代码与案例实践**

项目描述：基本上很多的互联网企业的主要收入来自于广告。其中对收入影响最大的就是广告的点击率，而且需要根据预测的点击率来优化广告列表的展示。

- 数据预处理
- 特征工程
- 选择合适的特征
- 逻辑回归模型的搭建
- 利用正则去避免过拟合现象
- 模型的评估

**项目学习目标**

- 掌握特征工程的技术
- 掌握机器学习建模的整套流程
- 深入理解逻辑回归模型以及设计原理
- 深入理解过拟合现象，以及如何在实际应用中避免此类现象

第六节：情感分析项目（4 学时）

- 情感分析技术概览
- 朴素贝叶斯介绍
- 支持向量机介绍
- 文本预处理技术介绍
- tf-idf 以及语言模型
- 词向量技术
- ELMo, Bert 介绍
- 词性分析
- 样本不均衡的处理

代码与案例实践

项目描述：在文本分析领域，情感分析是应用场景最为广泛的技术。在此项目中，我们结合分类算法以及各类特征工程技术来搭建基于评论的情感分析系统。

- 文本数据的读取
- 文本的预处理
- 把文本表示成向量
- 提取文本的特征
- 搭建分类模型
- 利用各类技术来调参
- 利用 F1-score 来评估系统

项目学习目标

- 掌握文本分析流程
- 掌握分类模型的使用
- 掌握文本特征工程

第 4 天 | 第七节：金融风控预测项目（4 学时）

- 评分卡介绍
- 信息熵与决策树模型
- 决策树的构建流程详解
- 随机森林模型的介绍
- Bootstrap 介绍
- 集成模型的优势
- GBDT 介绍
- GBDT 与随机森林的区别
- XGBoost 介绍

代码与案例实践

项目描述：在金融科技领域，风控是核心。在此项目中，我们需要搭建一个风控系统，此系统可以帮助我们更好地管理用户的风险，从而更准确地做放贷与否的决策。

- |         |           |
|---------|-----------|
| • 数据的清洗 | • 集成模型的搭建 |
| • 特征工程  | • 搭建分类模型  |
| • 特征的转换 | • 集成模型的评估 |

### 项目学习目标

- 掌握如何做特征工程以及特征的转换
- 深入理解怎么使用各类的集成学习模型

## 第八节：推荐系统项目（4 学时）

- 推荐技术的概览
- 协同过滤技术
- 矩阵分解技术详解
- 利用 SGD 求解矩阵分解
- Cold-Start 问题的处理
- 考虑 side information
- 考虑时间的矩阵分解
- 主题模型介绍
- 吉布斯采样
- 利用主题模型抽取文本特征

### 代码与案例实践

项目描述：推荐系统已经成为了绝大部分互联网产品的核心，根据用户的兴趣推荐最合适的产品来提升营收。在此项目中，我们利用文本数据来搭建新闻推荐系统。

- 数据的清洗
- 数据预处理
- 矩阵分解求解隐式向量
- 结合文本信息
- 推荐模型的评估

### 项目学习目标

- 掌握如何使用矩阵分解的技术来做推荐
- 掌握如何使用额外的特征信息来提升推荐的准确率



## 第 5 天 | 第九节：图像识别项目（4 学时）

- 神经网络模型介绍
- 深度神经网络的优势
- 理想中的表示学习
- 梯度消减现象
- Keras 的安装与使用
- 安装 GPU 环境
- 卷积神经网络介绍
- 卷积层, Pooling 层介绍
- Dropout 介绍
- 经典的 CNN 结构
- Batch Normalization

### 代码与案例实践

项目描述：在本项目中，我们利用卷积神经网络模型来实现交通牌的识别。由于噪声的影响，这个问题拥有很大的挑战。

- 图像数据的预处理
- 图像增强
- 搭建卷积神经网络模型
- 在 GPU 环境下训练模型
- 调参

### 项目学习目标

- 掌握使用 CNN 模型来识别图像
- 掌握使用图像增强技术
- 掌握图片去噪

## 第十节：机器翻译项目（4 学时）

- 机器翻译技术概览
- RNN 模型介绍
- RNN 的梯度爆炸
- LSTM 与 GRU 模型介绍
- Seq2Seq 模型介绍
- Attention 与 Self-Attention
- Transformer 与 Bert
- 训练过程
- 机器翻译系统评估

### 代码与案例实践

项目描述：在本项目中我们利用 Seq2Seq 模型来搭建中英机器翻译系统，并在 GPU 环境下运行。

- 文本的预处理
- Seq2Seq 模型的搭建
- 在 GPU 环境下训练
- 调参
- Debug

### 项目学习目标

- 掌握如何 Seq2Seq 模型来搭建机器翻译系统。

## 第 6 天 | 第十一节：风控知识图谱项目（8 学时）

- 知识图谱介绍
- 图数据库介绍 :Neo4j
- 实体与关系
- 知识图谱的设计
- 命名实体识别
- 实体消歧
- 实体统一
- 图算法介绍
- GraphX 的使用

### 代码与案例实践

项目描述：知识图谱已经成为人工智能领域关键技术之一。在本项目中，我们搭建用于做风控的知识图谱。

- 数据清洗
- 文本的预处理
- 设计知识图谱
- 编写图算法
- 评估系统

### 项目学习目标

- 掌握如何使用知识图谱技术来搭建风控系统

## 第十二节：总结以及未来展望