

# 课程大纲

## 壹、 培训内容：

模块一：详解 machine learning 算法及应用

模块二：企业机器学习平台搭建

模块三：如何解决小数据问题：

模块四：如何进行算法优化

## 贰、 课程时间：4 天，6 小时/天

## 参、 授课对象：

## 肆、 授课方式：

理论讲授，案例分析，方法传授、动画演示、互动讨论，讲师点评、实战演练。

## 伍、 课程大纲：

# 模块一：详解 machine learning 算法及应用

## 1. 实用算法理论

### 1.1 信息检索 (NER 关系抽取)

#### 1.1.1 引言

#### 1.1.2 NN/CNN-CRF 模型

#### 1.1.3 RNN-CRF 模型

#### 1.1.4 最近的一些工作

#### 1.1.5 总结

### 1.2 文本摘要

#### 1.2.1 介绍

#### 1.2.2 生成式文本摘要

#### 1.2.3 最新前沿动态

#### 1.2.4 评估摘要

#### 1.2.5 文本摘要总结

### **1.3 多类分类与多标签分类**

#### **1.3.1 多类与多标签分类两者区别**

#### **1.3.2 性能度量方法**

#### **1.3.3 基于样例的度量方法**

#### **1.3.4 基于标签的度量方法**

#### **1.3.5 深度学习的应用**

### **1.4 OCR**

#### **1.4.1 OCR 简介**

#### **1.4.2 OCR 的应用场景**

#### **1.4.3 OCR 的技术路线**

#### **1.4.4 图像预处理**

#### **1.4.5 文字检测**

#### **1.4.6 文本识别**

#### **1.4.7 端到端的 OCR**

## **2. 实用算法应用**

### **2.1 智能问答对话的架构及各个组件（多轮对话，知识图谱、业务问答、闲聊）**

### **2.2 搜索引擎**

### **2.3 用户行为分析（多模态）**

### **2.4 图表检索（柱状图等）**

## **模块二：企业机器学习平台搭建**

### **1. Anaconda 在线平台搭建**

#### **1.1 python 环境安装**

#### **1.2 Numpy、Pandas 等数据操作环境的安装**

#### **1.3 Tensorflow 环境安装**

#### **1.4 Keras 环境安装**

#### **1.5 Sklearn 机器学习环境的安装**

#### **1.6 在线并发环境编程实践**

## **模块三：如何解决小数据问题：**

### **2. 数据标注系统 - 用户**

- 2.1 开源的数据源介绍 (10 种)
- 2.2 数据标注系统及实践 (2 种)
- 3. **主动学习 (active learning) 模型**
  - 3.1 主动学习模型结构
  - 3.2 查询函数举例
  - 3.3 主动学习编程实践
- 4. **迁移学习模型**
  - 4.1 迁移学习思路
  - 4.2 迁移学习编程实践 (vgg16)
- 5. **无监督学习 (聚类算法)**
  - 5.1 聚类算法
  - 5.2 聚类算法编程实践
- 6. **增强学习 (reinforcement learning)**
- 7. **数据增强**
  - 7.1 数据增强方法介绍 (5 种)
  - 7.2 数据增强编程实践

## 模块四：如何进行算法优化

- 1. 预处理与特征工程
- 2. 神经网络调参之路
- 3. 模型融合

### 六、课程总结：

- 一、重点知识回顾与总结
- 二、互动与讨论：问与答

就学员提出的问题进行分析、讨论、模拟演练和点评。

### 七、讲师介绍：