



附件：BDA 大数据分析师培训大纲

主讲：王老师

报名电子邮件：

Liang1994bj@126.com

培训地点：北京

培训时间：2 天

课程定制：可根据用户需求调整课程内容。

## 自然语言处理解决方案实战

### 培训大纲

#### 课程概述

本次课程面向对大数据分析、挖掘等感兴趣的研究人员，带大家实践大数据分析挖掘与建模系统的项目训练，成体系地讲解：

- (1) 大数据建模、分析挖掘技术体系及其平台方案实现；
- (2) 大数据支持的自然语言处理技术-方法与实际应用案例讲解；

#### 设备要求

学员需要准备的笔记本配置 i5 及以上 CPU，4GB 及以上内存，硬盘空间预留 50GB（可用移动硬盘）。

#### 培训参考教材

《数据挖掘导论（完整版）》

《Python 自然语言处理》



# 中国信息协会市场研究业分会

## China Marketing Research Association

---

### 课程安排

---

#### 第一天 机器学习与 AI 方法：典型应用场景与新进展

1. 大数据颠覆了传统的理念和技术吗?
    - (1) 大数据确实带来了新的理念和技术体系
    - (2) 也推动了一些新算法的诞生
    - (3) 那么，大数据时代，传统理念还有用吗？
    - (4) 传统技术还有用吗？
  2. 常见分类模型
    - (1) Bagging
    - (2) 随机森林 RF
    - (3) SVM 和 RF
    - (4) Adaboost
    - (5) 近几年 GBDT\XGBOOST\LightGBM 渐成流行
    - (6) 逻辑斯特回归 (LR)
    - (7) KNN
    - (8) 朴素贝叶斯
  3. 常见聚类模型
    - (1) K 均值聚类
    - (2) 均值漂移算法
    - (3) 基于密度的聚类算法 (DBSCAN)
    - (4) 利用高斯混合模型进行最大期望估计
    - (5) 凝聚层次聚类
  4. 常见回归模型
    - (1) 线性回归
    - (2) 多项式回归
-



# 中国信息协会市场研究业分会

China Marketing Research Association

---

---

(3) Ridge 回归

(4) Lasso 回归

(5) ElasticNet 回归

## 5. 常见关联模型

(1) 关联分析的基本概念

(2) 发现频繁项集

(3) 发现关联规则

## 6. 有监督、无监督和半监督

(1) 有监督学习

(2) 无监督学习

(3) 半监督学习

## 7. 认识 NLP

(1) 认识 NLP

(2) NLP 对人工智能的意义

(3) NLP 基础--词特征

(4) NLP 基础--词向量

(5) 词干提取

(6) 词形还原

(7) 词性标注

(8) 命名实体消歧

(9) 命名实体识别

(10) 自然语言推理

## 第二天 NLP 实战--典型场景与解决方案

### 1. 舆情监控

(1) 场景还原

(2) 解决方案

---



# 中国信息协会市场研究业分会

China Marketing Research Association

---

---

(3) 代码讲解

(4) 上机操作

(5) 操作答疑

## 2. 文本摘要

(1) 场景还原

(2) 解决方案

(3) 代码讲解

(4) 上机操作

(5) 操作答疑

## 3. 语义相似

(1) 场景还原

(2) 解决方案

(3) 代码讲解

(4) 上机操作

(5) 操作答疑

---

### 专家介绍

王老师：高级统计师、经济学博士、数据挖掘方向博士后、北京大学客座教授，现为某大型金融集团 AI 部门数据挖掘团队负责人，曾在国家统计局工作 10 年，研究方向为金融与医疗大数据挖掘、统计建模、机器学习与深度学习算法，在大数据应用、人工智能领域有丰富实战项目经验。