



子雨大数据
BIG DATA

《人工智能通识》

教材官网: <https://dblab.xmu.edu.cn/post/ai-introduction/>

第4章 自然语言处理 备课方法

(PPT版本号: 2026年春季学期)



访问教材官网

林子雨 博士/副教授

厦门大学计算机科学与技术系

E-mail: ziyulin@xmu.edu.cn ▶▶

主页: <http://dblab.xmu.edu.cn/post/linziyu>





提纲

- 教学目标
- 教学重点
- 教学难点
- 学情分析
- 教学方法设计
- 教学内容与授课节奏

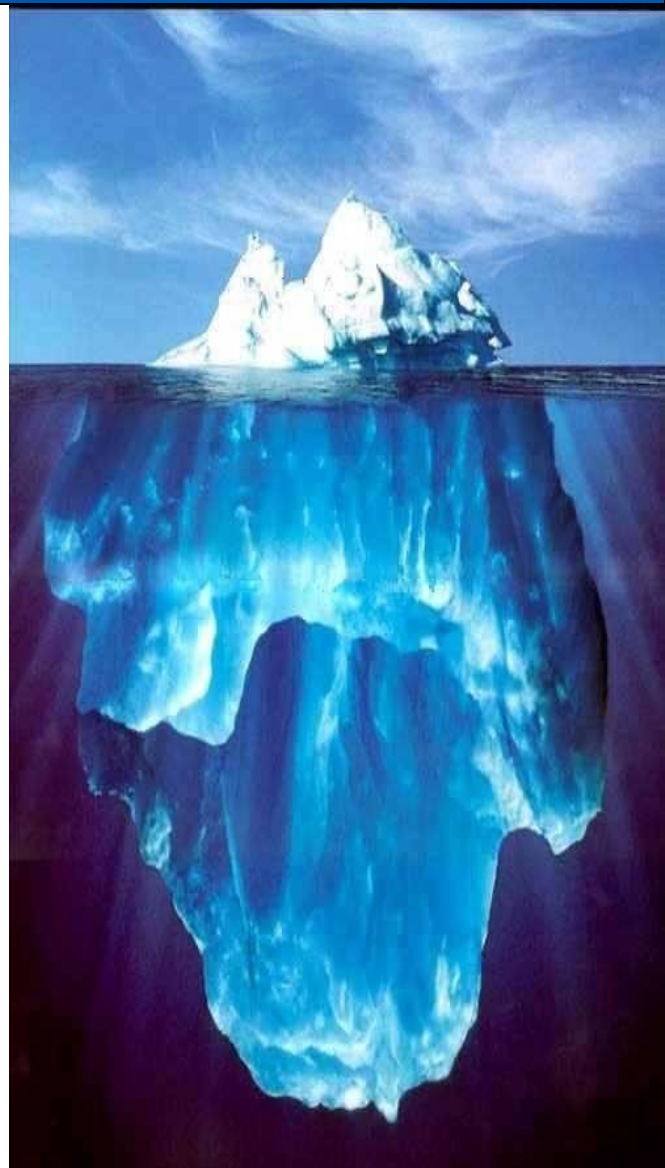


高校大数据课程

公共服务平台



<http://dbleab.xmu.edu.cn/post/bigdata-teaching-platform/>





章节整体定位

1.本章定位

- 自然语言处理（NLP）是人工智能最贴近人类、最常用、最易感知的技术方向，是学生理解“AI能听懂、会说话、会写作”的关键一章

2.教学原则

- 不讲复杂公式、不堆语言学理论
- 不讲模型推导、不要求底层原理
- 只讲：是什么→干什么→怎么用→生活案例

3.教学目标人群：文理工医艺体全覆盖，零编程、零语言学基础也能听懂



教学目标

1. 知识目标

- 理解自然语言处理的定义、目标、价值
- 了解 **NLP** 四个发展阶段（规则→统计→词向量→深度学习 / 大模型）
- 掌握**5** 大核心基础任务：文本分类、关系抽取、命名实体识别、文本聚类、信息检索
- 熟悉**5** 大典型应用：机器翻译、智能问答、文本摘要、情感分析、对话机器人
- 了解两个入门实验：**Jieba** 分词、情感分析



教学目标

2. 能力目标

- 能识别生活中的 **NLP** 应用并说出对应技术
- 能区分 **NLP**基础任务与上层应用的关系
- 能看懂简单文本处理流程：分词→预处理→分析→输出
- 能运行极简 **NLP** 代码，理解“机器处理文字”的过程



教学目标

3. 素养目标

- 理解语言是人类智慧核心，NLP 是AI认知智能的关键
- 建立对大模型、聊天机器人的理性认知
- 认识技术伦理：偏见、隐私、误导信息等问题



教学重点

- NLP 的定义与核心价值
- NLP发展四阶段（帮助学生建立技术演进观）
- 5 大核心基础任务（理解 NLP 底层能力）
- 5 大典型应用场景（与生活强绑定）
- 两个轻量化实验：Jieba 分词、情感分析



教学难点

- 学生容易混淆：NLP 基础任务 vs 上层应用
- 难以理解：语言为什么难处理（歧义、省略、隐喻、语境）
- 抽象概念：词嵌入 / 语义表示（不讲数学，只讲“把词变成向量”）
- 区分：文本分类（有标签）vs 文本聚类（无标签）
- 理解：大模型与 NLP 的关系



学情分析

- 全员零基础：不懂语法树、不懂神经网络、不懂机器学习
- 对“语言”习以为常：不理解计算机处理文字有多难
- 只关心应用：ChatGPT、翻译、语音转文字、智能客服
- 喜欢可视化、案例、互动
- 代码恐惧：实验只看结果，不要求编写



教学方法设计

1. 案例导入法（贯穿全章）
 - 机器翻译→打破语言壁垒
 - 情感分析→电商评论、舆情监控
 - 智能问答→小爱同学、Siri
 - 文本摘要→新闻速读、文献总结
 - 对话机器人→ChatGPT、客服机器人
2. 对比教学法
 - 人类理解语言 vs 机器理解语言
 - 文本分类（有答案）vs 文本聚类（无答案）
 - 规则方法 vs 深度学习方法
3. 可视化教学法
 - 分词结果展示
 - 词云图可视化



教学内容与授课节奏

第 1 节 自然语言处理概述

(1) 什么是 NLP

一句话讲清：让计算机听懂、读懂、会说、会写人类语言的技术

目标：实现人与机器自然语言沟通

(2) NLP 的四层价值

- 技术层：AI 认知智能的核心
- 产业层：改造教育、医疗、金融、内容创作
- 社会层：跨语言沟通、信息平权、无障碍辅助
- 个人层：生活更方便（语音、翻译、问答）

(3) 发展简史（四阶段，必讲）

- 规则阶段：人写语法
- 统计阶段：从语料库学规律
- 词向量阶段：把词变成数字向量
- 深度学习 / 大模型：Transformer、BERT、GPT、通用理解



教学内容与授课节奏

第 2 节 自然语言处理核心基础任务

（本章最干货，不讲原理，只讲功能 + 用途）

1. 文本分类（情感分析、主题识别）

- 功能：给文本打标签
- 例子：评论正面 / 负面、新闻属于体育 / 科技
- 关键词：有监督、打标签、情感、主题

2. 关系抽取

- 功能：找出“谁和谁是什么关系”
- 例子：马斯克 —CEO→特斯拉
- 输出：实体 - 关系 - 实体三元组



教学内容与授课节奏

3. 命名实体识别 (NER)

- 功能：抓出文本关键信息
- 例子：人名、地名、机构、时间
- 地位：所有 NLP 系统的入口

4. 文本聚类

- 功能：把相似文本自动归组
- 对比：分类 = 有答案；聚类 = 无答案
- 例子：评论自动分堆、新闻自动聚类

5. 信息检索

- 功能：快速找到相关内容
- 例子：百度搜索、文献检索
- 目标：相关、快速、准确

讲法：每个任务 3 句话：功能→例子→用途



教学内容与授课节奏

第 3 节 典型应用场景

（学生最爱听，全用生活案例）

- 机器翻译：百度翻译、讯飞翻译机
- 智能问答：小爱、小度、知识库问答
- 文本摘要：长文变短、新闻速读
- 情感分析：电商评论、舆情监控
- 对话系统：ChatGPT、智能客服

讲法：每一个应用都问学生：你用过吗？为什么好用？



教学内容与授课节奏

第 4 节 实验 1: Jieba 分词

核心讲法

- **Jieba:** 中文分词“标配工具”
- 三种模式: 精确、全、搜索引擎
- 实验目标: 看到文字被切开
- 词云图: 直观看到关键词

通识课要求: 教师会运行, 学生看结果, 不要求写代码



教学内容与授课节奏

第5 节 实验 2：情感分析

核心讲法

- 输入一句话 → 输出正面 / 负面 / 中性
- 工具：TextBlob
- 目标：理解“机器能读懂情绪”

The background features several faint, light-blue silhouettes of people. At the top, there are two groups of people, one on the left and one on the right, appearing to be in conversation or holding hands. On the right side, there is a larger silhouette of a person standing with their hand to their face. In the bottom left corner, there are silhouettes of two people sitting at a table, possibly in a meeting or classroom setting. The overall scene is set against a dark blue gradient background.

Thank You!

Department of Computer Science, Xiamen University, 2026