

## Chapter1: 绪论

- 基本概念
- 研究内容
- 基本问题
- 面临困难
- 基本方法

## Chapter2: 数学基础

- 最大似然估计 MLE
- 全概率公式
- 贝叶斯定理
- 熵
  - 联合熵
  - 条件熵
  - 相对熵
  - 交叉熵
- 困惑度
- 互信息
- 双字耦合度

## Chapter4: 语料库与语言知识库

- 语料库
  - 宾夕法尼亚大学 Tree Bank
  - 北京大学 CLKB
  - 口语语料库 BTEC
- 词汇知识库
  - 普林斯顿大学 WordNet
- 知网
- 语义关系

## Chapter5: 语言模型

- 传统语言模型
  - n 元文法
  - 参数估计
  - 数据平滑
  - 语言模型自适应
- 神经网络模型
  - 词向量
  - 前馈神经网络 FNN
  - 循环神经网络 RNN
    - 长短时记忆网络 LSTM
  - 自注意力机制
- Softmax 回归
- 文本表示

word2vec

CBOW 模型

Skip-gram 模型

## Chapter6: 马尔科夫模型和条件随机场

马尔科夫模型

隐马尔科夫模型 HMM

三个问题

前向算法

后向算法

维特比 Viterbi 搜索算法

参数学习

条件随机场 CRFs

## Chapter7: 自动分词、命名实体识别与词性标注

汉语自动分词

歧义切分处理

分词性能评价

正确率 P

召回率 R

分词基本算法

最大匹配法 MM

最少分词法

基于语言模型

基于序列标注

未登录词识别

命名实体识别

词性标注

## Chapter8: 句法分析

线图分析法

CYK 分析法

基于 PCFG 的分析法

句法分析性能评估

局部句法分析

## Chapter9: 语义分析

贝叶斯消歧

基于语义定义的消歧

基于义类辞典的消歧

## Chapter13: 文本分类与聚类

文本分类

文本表示

向量空间模型 VSM (词袋模型 BOW)

特征选择

文档频率 DF

互信息 MI

信息增益 IG

Chi-Square 统计

分类算法

朴素贝叶斯

线性支持向量机 SVM

最邻近算法 KNN

文本聚类

K-means 算法

K-medoids 算法

Chapter14: 深度神经网络

卷积神经网络 CNN

循环神经网络 RNN