

文本数据分析

Table of Contents

第一章 文本数据基础	1
1.1 什么是文本数据	1
1.2 文本数据的常见来源	1
第二章 文本预处理	3
2.1 分词	3
2.2 去除停用词	3
2.3 大小写转换	3
2.4 标点符号处理	3
2.5 词干提取与词形还原	3
第三章 特征提取	5
3.1 词袋模型 (Bag of Words)	5
3.2 TF-IDF	5
3.3 词嵌入 (Word Embedding)	5
3.4 文本分析方法	5
3.5 词频统计	5
3.6 主题建模 (如 LDA)	5
3.7 情感分析	5
3.8 文本分类	5
第四章 可视化	7
4.1 词云	7
4.2 频率分布图	7

第五章 实践案例	9
5.1 新闻文本分析	9
5.2 评论情感分析	9

第一章 文本数据基础

1.1 什么是文本数据

1.2 文本数据的常见来源

第二章 文本预处理

2.1 分词

2.2 去除停用词

2.3 大小写转换

2.4 标点符号处理

2.5 词干提取与词形还原

第三章 特征提取

3.1 词袋模型 (Bag of Words)

3.2 TF-IDF

3.3 词嵌入 (Word Embedding)

3.4 文本分析方法

3.5 词频统计

3.6 主题建模 (如 LDA)

3.7 情感分析

3.8 文本分类

第四章 可视化

4.1 词云

4.2 频率分布图

第五章 实践案例

5.1 新闻文本分析

5.2 评论情感分析

