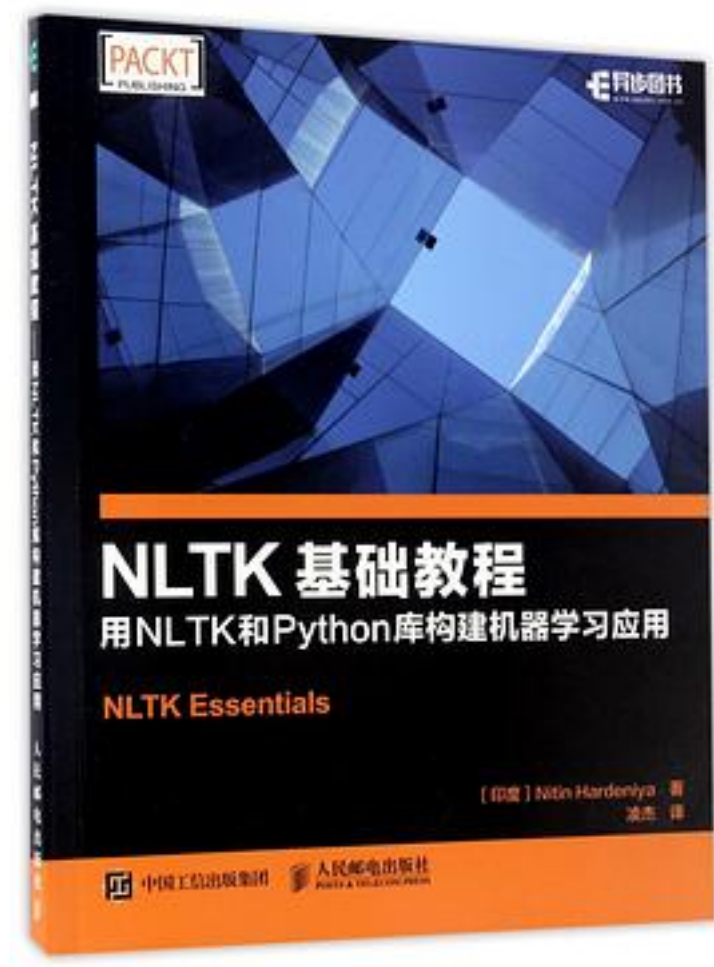


NLTK基础教程(用NLTK和Python库构建机器学习应用)



[NLTK基础教程\(用NLTK和Python库构建机器学习应用\)_下载链接1](#)

著者:[印度]哈登尼亚 (Nitin Hardeniya)

出版者:人民邮电出版社

出版时间:2017-6

装帧:平装

isbn:9787115452573

本书主要介绍如何通过NLTK库与一些Python库的结合从而实现复杂的NLP任务和机器学习应用。全书共分为10章。第1章对NLP进行了简单介绍。第2章、第3章和第4章主要介绍一些通用的预处理技术、专属于NLP领域的预处理技术以及命名实体识别技术等。第5章之后的内容侧重于介绍如何构建一些NLP应用，涉及文本分类、数据科学和数据处理、社交媒体挖掘和大规模文本挖掘等方面。

本书适合 NLP

和机器学习领域的爱好者、对文本处理感兴趣的读者、想要快速学习NLTK的zishenPython程序员以及机器学习领域的研究人员阅读。

作者介绍:

Nitin Hardeniya

数据科学家，拥有4年以上从业经验，期间分别任职于Fidelity、Groupon和[24]7等公司，其业务横跨各个不同的领域。此外，他还拥有IIT-H的计算语言学硕士学位，并且是5项客户体验专利的作者。

目录: 第1章自然语言处理简介1

1.1为什么要学习NLP2

1.2先从Python开始吧5

1.2.1列表5

1.2.2自助功能6

1.2.3正则表达式8

1.2.4字典9

1.2.5编写函数10

1.3向NLTK迈进11

1.4练习16

1.5小结17

第2章文本的歧义及其清理18

2.1何谓文本歧义18

2.2文本清理20

2.3语句分离器21

2.4标识化处理22

2.5词干提取23

2.6词形还原24

2.7停用词移除25

2.8罕见词移除26

2.9拼写纠错26

2.10练习27

2.11小结28

第3章词性标注29

3.1何谓词性标注29

3.1.1Stanford标注器32

3.1.2深入了解标注器33

3.1.3顺序性标注器35

3.1.4Brill标注器37

3.1.5基于机器学习的标注器37

3.2命名实体识别（NER）38

3.3练习40

3.4小结41

第4章文本结构解析43

4.1浅解析与深解析43

4.2两种解析方法	44
4.3为什么需要进行解析	44
4.4不同的解析器类型	46
4.4.1递归下降解析器	46
4.4.2移位—归约解析器	46
4.4.3图表解析器	46
4.4.4正则表达式解析器	47
4.5依存性文本解析	48
4.6语块分解	50
4.7信息提取	53
4.7.1命名实体识别 (NER)	53
4.7.2关系提取	54
4.8小结	55
第5章NLP应用	56
5.1构建第一个NLP应用	57
5.2其他NLP应用	60
5.2.1机器翻译	60
5.2.2统计型机器翻译	61
5.2.3信息检索	62
5.2.4语音识别	64
5.2.5文本分类	65
5.2.6信息提取	66
5.2.7问答系统	67
5.2.8对话系统	67
5.2.9词义消歧	67
5.2.10主题建模	68
5.2.11语言检测	68
5.2.12光符识别	68
5.3小结	68
第6章文本分类	70
6.1机器学习	71
6.2文本分类	72
6.3取样操作	74
6.3.1朴素贝叶斯法	76
6.3.2决策树	79
6.3.3随机梯度下降法	80
6.3.4逻辑回归	81
6.3.5支持向量机	81
6.4随机森林算法	83
6.5文本聚类	83
6.6文本中的主题建模	84
6.7参考资料	87
6.8小结	87
第7章Web爬虫	88
7.1Web爬虫	88
7.2编写第一个爬虫程序	89
7.3Scrapy库中的数据流	92
7.3.1Scrapy库的shell	93
7.3.2目标项	98
7.4生成网站地图的蜘蛛程序	99
7.5目标项管道	100
7.6参考资料	102
7.7小结	102
第8章NLTK与其他Python库的搭配运用	104

8.1NumPy104	
8.1.1多维数组105	
8.1.2基本运算106	
8.1.3从数组中提取数据107	
8.1.4复杂矩阵运算108	
8.2SciPy112	
8.2.1线性代数113	
8.2.2特征值与特征向量113	
8.2.3稀疏矩阵114	
8.2.4优化措施115	
8.3pandas117	
8.3.1读取数据117	
8.3.2数列119	
8.3.3列转换121	
8.3.4噪声数据121	
8.4matplotlib123	
8.4.1子图绘制123	
8.4.2添加坐标轴124	
8.4.3散点图绘制125	
8.4.4条形图绘制126	
8.4.53D绘图126	
8.5参考资料126	
8.6小结127	
第9章Python中的社交媒体挖掘128	
9.1数据收集128	
9.2数据提取132	
9.3地理可视化134	
9.3.1影响力检测135	
9.3.2Facebook135	
9.3.3有影响力的朋友139	
9.4小结141	
第10章大规模文本挖掘142	
10.1在Hadoop上使用Python的不同方式142	
10.1.1Python的流操作143	
10.1.2Hive／Pig下的UDF143	
10.1.3流封装器143	
10.2Hadoop上的NLTK144	
10.2.1用户定义函数（UDF） 144	
10.2.2Python的流操作146	
10.3Hadoop上的Scikit—learn147	
10.4PySpark150	
10.5小结153	
• • • • •	(收起)

[NLTK基础教程\(用NLTK和Python库构建机器学习应用\)_ 下载链接1_](#)

标签

机器学习

NLP

python

Python

计算机科学

计算机

nltk

数据分析

评论

入门的好教程，可以当作大纲来看，同时也说明这本书讲得不够深入。但是，扣分的点在于书中有些代码跑不起来，不够严谨

代码好多错误，跑不起来

流程介绍简洁易懂 知识面广 不错的入门书。需要再看第二遍

从这本书里学到了如何使用朴素贝叶斯对文本进行分类，对我分析微博特别有帮助，整体上而言是非常好的一本书，内容非常丰富，自然语言和数据挖掘都能覆盖到，强烈推荐。

[NLTK基础教程\(用NLTK和Python库构建机器学习应用\) 下载链接1](#)

书评

- 1、版本较老，书内Python版本为2.6.6，所以对应的nltk版本可能也很老。
- 2、全书都是讲的英文处理，并未涉及到任何中文内容。NLP的中文和英文有很大差别，所以这本书对中文NLP的指导性并不高。
- 1、版本较老，书内Python版本为2.6.6，所以对应的nltk版本可能也很老。
- 2、全书都是...

[NLTK基础教程\(用NLTK和Python库构建机器学习应用\)_下载链接1](#)