

《大数据时代的算法：机器学习、人埂

图书基本信息

书名：《大数据时代的算法：机器学习、人工智能及其典型实例》

13位ISBN编号：9787121304295

出版时间：2017-1

作者：刘凡平

页数：220

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu111.com

《大数据时代的算法：机器学习、人埂

内容概要

《大数据时代的算法：机器学习、人工智能及其典型实例》介绍在互联网行业中经常涉及的算法，包括排序算法、查找算法、资源分配算法、路径分析算法、相似度分析算法，以及与机器学习相关的算法，包括数据分类算法、聚类算法、预测与估算算法、决策算法、关联规则分析算法及推荐算法。《大数据时代的算法：机器学习、人工智能及其典型实例》涉及的相关算法均为解决实际问题中的主流算法，对于工作和学习都有实际参考意义。

《大数据时代的算法：机器学习、人工智能及其典型实例》是一本算法领域内的技术参考书籍，涵盖数十种算法，通过由浅入深的介绍基础算法和机器学习算法相关理论和应用，阐述了各个算法的应用场景及算法复杂度，使读者对算法的理解不只是停留在表面，还从应用的角度提供了大量实例，使读者能够快速、高效进阶各类算法，并能够熟练应用到将来的工作实践中。

书籍目录

第1章 算法基础

1

1.1 基础算法分析类型

1

1.1.1 分治法

1

1.1.2 动态规划法

2

1.1.3 回溯法

3

1.1.4 分支限界法

4

1.1.5 贪心法

4

1.2 算法性能分析

5

1.3 概率论与数理统计基础

6

1.4 距离计算

8

1.4.1 欧氏距离

8

1.4.2 马氏距离

9

1.4.3 曼哈顿距离

9

1.4.4 切比雪夫距离

9

1.4.5 闵氏距离

9

1.4.6 海明距离

10

1.5 排序算法

10

1.5.1 快速排序

11

1.5.2 归并排序

11

1.5.3 堆排序

13

1.5.4 基数排序

15

1.5.5 外排序

16

1.6 字符压缩编码

17

1.6.1 哈夫曼编码

17	
1.6.2	香农-范诺编码
21	
1.7	本章小结
24	
第2章	数据查找与资源分配算法
25	
2.1	数值查找算法
25	
2.1.1	二分搜索算法
25	
2.1.2	分块查找算法
27	
2.1.3	哈希查找算法
28	
2.2	字符串查找算法
30	
2.2.1	Knuth-Morris-Pratt算法
31	
2.2.2	Boyer-Moore算法
34	
2.2.3	Sunday算法
37	
2.3	海量数据中的查找
39	
2.3.1	基于布隆过滤器查找
39	
2.3.2	倒排索引查找
41	
2.4	银行家算法
43	
2.5	背包问题
44	
2.5.1	0-1背包问题
45	
2.5.2	部分背包问题
46	
2.6	本章小结
47	
第3章	路径分析算法
49	
3.1	基于Dijkstra算法的路径分析
49	
3.1.1	应用示例：极地探险
49	
3.1.2	基于Dijkstra的最短路径规划
50	
3.2	基于Floyd算法的路径分析
53	

3.2.1 应用示例：任意两个城市之间的最短路径

53

3.2.2 Floyd原理

54

3.2.3 基于Floyd算法计算两个城市最短距离

56

3.3 基于A*算法的路径搜索

58

3.3.1 应用实例：绕过障碍区到达目的地

58

3.3.2 A*算法与最短距离计算

59

3.4 基于维特比算法的概率路径

61

3.4.1 应用实例：推断天气状态

61

3.4.2 维特比算法思想

62

3.4.3 计算天气状态

62

3.5 最长公共子序列问题

64

3.5.1 概要

64

3.5.2 最长公共子串

64

3.5.3 最长公共子序列原理

66

3.5.4 实例：求两字符串的最长公共子序列

66

3.6 本章小结

68

第4章 相似度分析算法

69

4.1 应用实例：海量网页相似度分析

69

4.2 基于Jaccard相似系数的相似度计算

70

4.2.1 计算流程

70

4.2.2 狭义Jaccard相似系数

71

4.2.3 广义Jaccard相似系数

71

4.3 基于MinHash的相似性算法

71

4.3.1 与Jaccard相似性关系

71

4.3.2 计算网页文本相似性过程

72	
4.4	向量空间模型
73	
4.4.1	词袋模型
73	
4.4.2	TF-IDF算法
74	
4.5	基于余弦相似性算法的相似度分析
76	
4.5.1	原理基础
76	
4.5.2	公式解析
77	
4.5.3	计算网页文本相似性过程
77	
4.6	基于语义主题模型的相似度算法
78	
4.7	基于SimHash算法的指纹码
80	
4.7.1	SimHash引入
81	
4.7.2	SimHash的计算流程
81	
4.7.3	计算重复信息
83	
4.8	相似度算法的差异性
84	
4.9	本章小结
85	
第5章	数据分类算法
86	
5.1	基于朴素贝叶斯分类器
86	
5.1.1	有监督分类与无监督分类
87	
5.1.2	应用实例：识别车厘子与樱桃
88	
5.1.3	分类流程归纳
91	
5.1.4	应用扩展：垃圾邮件识别
92	
5.1.5	常用评价指标
96	
5.2	基于AdaBoost分类器
100	
5.2.1	AdaBoost概述
100	
5.2.2	AdaBoost算法具体流程
101	

5.2.3	AdaBoost算法的应用实例	102
5.2.4	AdaBoost算法的优点	105
5.3	基于支持向量机的分类器	105
5.3.1	线性可分与线性不可分	106
5.3.2	感知器	107
5.3.3	支持向量机	108
5.4	基于K邻近算法的分类器	109
5.4.1	应用实例：电影观众兴趣发现	109
5.4.2	核心思想	109
5.4.3	电影观众兴趣发现	110
5.5	本章小结	113
第6章	数据聚类算法	115
6.1	基于系统聚类法	115
6.1.1	概述	116
6.1.2	最短距离法	117
6.1.3	重心聚类法	119
6.1.4	动态聚类法	120
6.2	基于K-Means聚类算法	122
6.2.1	应用实例：新闻聚类	122
6.2.2	逻辑流程	123
6.2.3	实现新闻聚类分析	124
6.2.4	K-Means++	128
6.2.5	K-中心点聚类算法	129
6.2.6	ISODATA聚类算法	130
6.3	基于密度的DBSCAN算法	

131	
6.4	基于BIRCH算法的聚类分析
133	
6.4.1	聚类特征
133	
6.4.2	聚类特征树
134	
6.5	聚类与分类差异
135	
6.6	本章小结
136	
第7章	数据预测与估算算法
137	
7.1	产生式模型与判别式模型
137	
7.2	基于最大似然估计的预测
138	
7.3	基于线性回归的估算
140	
7.3.1	概要
140	
7.3.2	最小二乘法
141	
7.4	基于最大期望算法分析
143	
7.5	基于隐马尔科夫模型预测
144	
7.5.1	应用实例：高温天气与行为概率
144	
7.5.2	原理分析
145	
7.5.3	高温天气与行为概率
147	
7.6	基于条件随机场的序列预测
151	
7.6.1	应用实例
151	
7.6.2	原理分析
151	
7.6.3	条件随机场的优缺点
153	
7.7	本章小结
154	
第8章	数据决策分析算法
155	
8.1	基于ID3算法的决策分析
156	
8.1.1	信息量
156	

8.1.2	信息熵	156
8.1.3	信息增益	157
8.1.4	ID3算法流程	157
8.1.5	ID3算法的应用	157
8.2	基于C4.5算法的分类决策树	159
8.2.1	概要	159
8.2.2	应用实例	159
8.3	基于分类回归树的决策划分	161
8.3.1	概要	162
8.3.2	应用实例：决策划分	163
8.3.3	剪枝	164
8.4	基于随机森林的决策分类	168
8.4.1	随机森林的特点	169
8.4.2	随机森林的构造方法	169
8.4.3	应用实例：决定车厘子的售价层次	170
8.5	本章小结	172
第9章	数据关联规则分析算法	174
9.1	基于Apriori算法的关联项分析	174
9.1.1	应用实例：超市的货架摆放问题	175
9.1.2	基本概要	175
9.1.3	算法原理	176
9.1.4	有效摆放货架	176
9.2	基于FP-Growth算法的关联性分析	179
9.2.1	构建FP树	179
9.2.2	频繁项分析	

181	
9.2.3	与Apripri算法比较
184	
9.3	基于Eclat算法的频繁项集挖掘
184	
9.4	本章小结
185	
第10章	数据推荐算法
187	
10.1	概要
187	
10.1.1	推荐算法发展
188	
10.1.2	协同过滤推荐
189	
10.2	基于Item-Based协同过滤推荐
190	
10.2.1	Item-Based基本思想
190	
10.2.2	Slope One实例：基于评分推荐
190	
10.3	基于User-Based协同过滤推荐
193	
10.3.1	应用实例：根据人群的推荐
194	
10.3.2	User-Based与Item-Based对比
197	
10.4	基于潜在因子算法的推荐
198	
10.4.1	应用实例：新闻推荐
198	
10.4.2	流行度与推荐
200	
10.5	推荐算法与效果评价
201	
10.6	本章小结
203	

《大数据时代的算法：机器学习、人埂

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu111.com