

Broadview[®]
www.broadview.com.cn

**BIG
DATA**

底层讲解：详细讲解了算法的设计思路，
体会大师的思想

涵盖面广：囊括常用的53种算法，用以
解决各类问题

应用广泛：可用在数据挖掘、商务智能、
广告与商品推荐等多个领域

大数据时代的 算法

机器学习、人工智能 及其典型实例

刘凡平◎编著

大数据时代，
只有算法能洞悉数据的内在逻辑，
让数据产生商业价值！



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

大数据时代的 算法

机器学习、人工智能 及其典型实例

刘凡平◎编著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书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。本书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。

本书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。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

大数据时代的算法：机器学习、人工智能及其典型实例 / 刘凡平编著. —北京：电子工业出版社，2017.1

ISBN 978-7-121-30429-3

I. ①大… II. ①刘… III. ①机器学习—研究②人工智能—研究 IV. ①TP18

中国版本图书馆 CIP 数据核字（2016）第 284501 号

策划编辑：张月萍

责任编辑：张 慧

印 刷：三河市鑫金马印装有限公司

装 订：三河市鑫金马印装有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱

邮编：100036

开 本：787×980 1/16 印张：13.5

字数：330 千字

版 次：2017 年 1 月第 1 版

印 次：2017 年 1 月第 1 次印刷

定 价：49.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888

质量投诉请发邮件至 zltts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819 faq@phei.com.cn。

致 谢

本书的内容来自于我对日常学习和工作的总结，我要感谢本人曾经的导师于炯教授、叶勇教授以及在微软和百度公司工作时的前辈王明雨、何绍建、黄诚，正是你们对我的悉心指导，才能使我有能力和勇气写出这本书，无论你们在哪里，凡平永远心存感激。当然还有那些曾经一起学习同学、共事的朋友，你们曾经给予我很多无私的帮助，使我在和你们相处的过程中能够快速成长，感谢一路上有你们的陪伴，正是有你们，沿途的风景才格外美丽。

感谢英国的 Peter Boden 先生，虽然我们素未谋面，但是你却一直支持我们的团队去实现自己的梦想，并连续两年无偿支持我们在 Github 开源社区的项目。在开源项目中，我们实现了不少有一定难度的算法，如果没有你的支持，也许我们不能正常完成相关的研发任务，更不可能去挑战搜索引擎中的各类算法，尤其在人工智能领域的算法，对你的感激之情溢于言表。

感谢我的创业伙伴魏琪，当我提出一起研究人工智能技术时，你毫不犹豫地选择了和我一起艰苦创业。无论处在创业的何种困难期，你总是为团队加油鼓气，你踏实和认真的工作态度让我钦佩不已，感谢你为本书的技术细节的修改提出了真知灼见，还有和我一起工作的同学和朋友都对本书提出了改进意见，对此也表示深深的感谢。

感谢北京源智天下公司的吉老师及电子工业出版社的各位编辑，对本书的编写都给予了极大的帮助，对本书的出版也付出了辛勤的劳动。

衷心感谢我的家人，感谢你们在过去一年中对我的理解和支持，为我营造了一个良好的写作环境，并鼓励我坚持认真写作，使本书能够顺利完成。

本书编写过程中还得到了很多朋友的支持和帮助，限于篇幅，虽然不能一一对你们表示感谢，但是我对你们一样心存感激。

最后，感谢这个时代给予了每位有理想的人实现人生价值的机会！



前 言

中国在很早就开始了算法研究，如《周髀算经》《九章算术》这类最具历史的算法书籍，后来的唐宋元明清各历史朝代也出现了《一位算法》《算法绪论》《算法全书》《算法统宗》等一系列算法名著，算法已经成为各行各业的基础研究。

本书通过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。本书是一本算法领域内的技术手册，涵盖数十种算法，不仅能使读者深入了解各类算法的基本理论，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。

本书特色

本书不仅将目前工程应用中主流的基础算法和机器学习算法都做了详尽的介绍，还囊括了当前热门算法内容，如数据分类算法、聚类算法、推荐算法等。本书充分利用了最新算法的应用研究结果，通过实例为读者展现了清晰的算法应用，不拘泥于算法枯燥的理论，更多地从实用价值、工程价值的角度将算法知识呈现给读者。

本书中的算法可以广泛应用于各个领域，可以在自然语言处理研究、数据分析与挖掘、

商务智能、广告与商品推荐等领域中深入应用。作者秉承数据结合算法产生价值的理论体系，在介绍算法的同时与数据紧密关联，并结合多年实际工作经验，将算法的内容阐述得淋漓尽致。本书中的算法研究在当前甚至未来相当一段时间内都具有很大的实际意义。

本书结构

本书按照由浅入深、循序渐进的顺序对大数据时代的算法进行介绍。全书内容共分为两大部分，十个章节。第一部分主要是针对基础领域算法的介绍和应用，包含第 1~4 章；第二部分主要针对机器学习领域算法的理论认识和实例解析，包含第 5~10 章。全书各章的主要内容如下。

第 1 章 算法基础

本章从算法的分析类型，如分治法、动态规划法、回溯法、分支限界法、贪心法入手开始介绍算法内容，还分析了算法的性能，并介绍了概率论与数理统计基础部分的内容。同时，对算法中常用的距离计算算法、排序算法及字符串压缩编码也做了完整介绍。

第 2 章 数据查找与资源分配算法

本章以数据的查找和资源分配作为突破口，介绍了常用的数值查找算法，如二分搜索算法、分块查找及哈希查找算法。除此之外，还介绍了常见的字符串查找算法及在海量数据中的查找算法：布隆过滤器和倒排索引查找，介绍了资源分配算法，包括常用的银行家算法和背包问题的解决算法。

第 3 章 路径分析算法

本章主要介绍了路径分析算法，包括基于 Dijkstra 算法、Floyd 算法、A*算法的路径分析方法。除了介绍传统的路径分析算法外，还介绍了维特比算法在概率中的路径选择，以及最长公共子串、最长公共子序列问题的求解算法。整个内容涵盖了绝大部分的路径选择算法。

第 4 章 相似度分析算法

本章主要介绍了相似内容的分析理论和应用，从简单的 Jaccard 相似系数开始入手，逐步深入到基于 MinHash 的相似性算法以及向量空间模型，向量空间模型已经成为众多算法的基础理论。后续还深入介绍了基于余弦相似性算法和基于语义主题模型的语义相似度算法，以及基于 SimHash 的指纹码重复值验证算法。

第 5 章 数据分类算法

本章集中介绍了数据分类算法的解决方案，从简单易于理解的朴素贝叶斯模型开始，由浅入深地介绍了 AdaBoost 分类器及支持向量机，它们都是数据分类的有效解决方案，还对机器学习的相关基础知识做了概要介绍，最后介绍了 K 邻近算法在数据分类中的应用。

第 6 章 数据聚类算法

本章介绍了数据聚类的相关算法，其中，无监督的聚类算法目前是比较热门的研究领域。首先介绍了传统的基于系统聚类的方法；然后介绍了基于 K-Means 聚类算法及基于密度的 DBSCAN 算法；最后介绍了基于 BIRCH 算法的聚类分析，通过聚类特征及聚类特征实现数据聚类。

第 7 章 数据预测与估算算法

本章介绍了数据的预测和估算的算法体系和应用范例，从产生式模型和判别式模型入手介绍各类模型的方法论。首先介绍了基于最大似然估计的预测以及基于线性回归的估算、基于最大期望算法；然后介绍了基于隐马尔科夫模型模型预测；最后介绍了基于条件随机场的序列预测。

第 8 章 数据决策分析算法

本章对数据决策的分析方法做了详细介绍，主要围绕决策树的理论基础展开。首先介绍了基于 ID3 算法的决策分析，包括信息熵、信息增益等；然后介绍了基于 C4.5 算法的分类决策树及基于分类回归树的决策划分；最后介绍了基于随机森林的决策分类。介绍过程中包含了大量实例。

第 9 章 数据关联规则分析算法

本章主要介绍了关联规则分析方法的理论和实践。Apriori 算法作为最常用的关联规则分析算法已经被广泛应用到各个领域，本章也对 Apriori 算法进行了深入的介绍，并对和 Apriori 算法同等重要的 FP-Growth 算法也通过实例做了详细介绍。本章最后还介绍了利用倒排文件思想的 Eclat 算法。

第 10 章 数据与推荐算法

本章主要介绍了数据与推荐算法中的应用关系，推荐算法作为目前各行各业最热门的算法之一，已经应用非常广泛。本章介绍了基于物品本身属性关系的 Item-Based 协同过滤推荐算法，以及基于 User-Based 协同过滤推荐算法。除此之外，还介绍了基于流行度和潜

在因子的推荐算法，以及推荐算法的效果评估相关内容。

本书通过对数据与算法相关理论介绍和应用，将理论和实际应用结合，并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，而是由浅入深地将基础算法和机器学习算法成熟应用到各个领域，达到游刃有余的状态。但因为每个算法都存在一定的缺点，所以希望读者能充分了解、掌握各个算法，将算法的作用发挥到极致。

除此之外，读者也能够在这本书中快速、高效地从大量数据中找出所需要的数据或其他信息，这在大数据时代起到了非常重要的作用，给读者带来极大的便利。

读者对象

- 对基本算法和机器学习算法有兴趣的读者。
- 对数据分析和统计学有兴趣的读者。
- 对算法有研究的基础算法、机器学习工程师。
- 互联网行业的不同层次从业者。
- 软件或计算机专业的在校大学生。

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 | 3.5.4 | 实例：求两字符串的最长公共子序列 | 66 |
| 2.3.2 | 倒排索引查找..... | 41 | 3.6 | 本章小结 | 68 |
| 2.4 | 银行家算法 | 43 | 第 4 章 | 相似度分析算法 | 69 |
| 2.5 | 背包问题..... | 44 | 4.1 | 应用实例：海量网页相似度分析... .. | 69 |
| 2.5.1 | 0-1 背包问题 | 45 | 4.2 | 基于 Jaccard 相似系数的相似 度计算 | 70 |
| 2.5.2 | 部分背包问题..... | 46 | 4.2.1 | 计算流程..... | 70 |
| 2.6 | 本章小结..... | 47 | 4.2.2 | 狭义 Jaccard 相似系数 | 71 |
| 第 3 章 | 路径分析算法 | 49 | 4.2.3 | 广义 Jaccard 相似系数 | 71 |
| 3.1 | 基于 Dijkstra 算法的路径分析 | 49 | 4.3 | 基于 MinHash 的相似性算法..... | 71 |
| 3.1.1 | 应用示例：极地探险..... | 49 | 4.3.1 | 与 Jaccard 相似性关系 | 71 |
| 3.1.2 | 基于 Dijkstra 的最短路径规划 | 50 | 4.3.2 | 计算网页文本相似性过程... .. | 72 |
| 3.2 | 基于 Floyd 算法的路径分析..... | 53 | 4.4 | 向量空间模型 | 73 |
| 3.2.1 | 应用示例：任意两个城市之间的最短路径 | 53 | 4.4.1 | 词袋模型 | 73 |
| 3.2.2 | Floyd 原理 | 54 | 4.4.2 | TF-IDF 算法..... | 74 |
| 3.2.3 | 基于 Floyd 算法计算两个城市最短距离 | 56 | 4.5 | 基于余弦相似性算法的相似度 分析 | 76 |
| 3.3 | 基于 A*算法的路径搜索 | 58 | 4.5.1 | 原理基础..... | 76 |
| 3.3.1 | 应用实例：绕过障碍区 到达目的地..... | 58 | 4.5.2 | 公式解析..... | 77 |
| 3.3.2 | A*算法与最短距离计算... .. | 59 | 4.5.3 | 计算网页文本相似性过程... .. | 77 |
| 3.4 | 基于维特比算法的概率路径 | 61 | 4.6 | 基于语义主题模型的相似度 算法 | 78 |
| 3.4.1 | 应用实例：推断天气状态... .. | 61 | 4.7 | 基于 SimHash 算法的指纹码..... | 80 |
| 3.4.2 | 维特比算法思想..... | 62 | 4.7.1 | SimHash 引入 | 81 |
| 3.4.3 | 计算天气状态 | 62 | 4.7.2 | SimHash 的计算流程 | 81 |
| 3.5 | 最长公共子序列问题..... | 64 | 4.7.3 | 计算重复信息 | 83 |
| 3.5.1 | 概要..... | 64 | 4.8 | 相似度算法的差异性 | 84 |
| 3.5.2 | 最长公共子串 | 64 | 4.9 | 本章小结 | 85 |
| 3.5.3 | 最长公共子序列原理..... | 66 | | | |

| | | | |
|----------------------------------|------------|--------------------------------|------------|
| 第 5 章 数据分类算法 | 86 | 6.1.3 重心聚类法 | 119 |
| 5.1 基于朴素贝叶斯分类器 | 86 | 6.1.4 动态聚类法 | 120 |
| 5.1.1 有监督分类与无监督 分类 | 87 | 6.2 基于 K-Means 聚类算法 | 122 |
| 5.1.2 应用实例：识别车厘子 与樱桃 | 88 | 6.2.1 应用实例：新闻聚类 | 122 |
| 5.1.3 分类流程归纳 | 91 | 6.2.2 逻辑流程 | 123 |
| 5.1.4 应用扩展：垃圾邮件 识别 | 92 | 6.2.3 实现新闻聚类分析 | 124 |
| 5.1.5 常用评价指标 | 96 | 6.2.4 K-Means++ | 128 |
| 5.2 基于 AdaBoost 分类器 | 100 | 6.2.5 K-中心点聚类算法 | 129 |
| 5.2.1 AdaBoost 概述 | 100 | 6.2.6 ISODATA 聚类算法 | 130 |
| 5.2.2 AdaBoost 算法具体流程 .. | 101 | 6.3 基于密度的 DBSCAN 算法 | 131 |
| 5.2.3 AdaBoost 算法的应用 实例 | 102 | 6.4 基于 BIRCH 算法的聚类分析 .. | 133 |
| 5.2.4 AdaBoost 算法的优点 | 105 | 6.4.1 聚类特征 | 133 |
| 5.3 基于支持向量机的分类器 | 105 | 6.4.2 聚类特征树 | 134 |
| 5.3.1 线性可分与线性不可分 ... | 106 | 6.5 聚类与分类差异 | 135 |
| 5.3.2 感知器 | 107 | 6.6 本章小结 | 136 |
| 5.3.3 支持向量机 | 108 | 第 7 章 数据预测与估算算法 | 137 |
| 5.4 基于 K 邻近算法的分类器 | 109 | 7.1 产生式模型与判别式模型 | 137 |
| 5.4.1 应用实例：电影观众 兴趣发现 | 109 | 7.2 基于最大似然估计的预测 | 138 |
| 5.4.2 核心思想 | 109 | 7.3 基于线性回归的估算 | 140 |
| 5.4.3 电影观众兴趣发现 | 110 | 7.3.1 概要 | 140 |
| 5.5 本章小结 | 113 | 7.3.2 最小二乘法 | 141 |
| 第 6 章 数据聚类算法 | 115 | 7.4 基于最大期望算法分析 | 143 |
| 6.1 基于系统聚类法 | 115 | 7.5 基于隐马尔科夫模型预测 | 144 |
| 6.1.1 概述 | 116 | 7.5.1 应用实例：高温天气与 行为概率 | 144 |
| 6.1.2 最短距离法 | 117 | 7.5.2 原理分析 | 145 |
| 6.1.3 重心聚类法 | 119 | 7.5.3 高温天气与行为概率 | 147 |
| 6.1.4 动态聚类法 | 120 | 7.6 基于条件随机场的序列预测 | 151 |
| 6.2 基于 K-Means 聚类算法 | 122 | 7.6.1 应用实例 | 151 |
| 6.2.1 应用实例：新闻聚类 | 122 | 7.6.2 原理分析 | 151 |
| 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 | | |

| | | | | | |
|--------------|--------------------------|------------|---------------|---------------------------------|------------|
| 7.6.3 | 条件随机场的优缺点..... | 153 | 9.1.4 | 有效摆放货架..... | 176 |
| 7.7 | 本章小结..... | 154 | 9.2 | 基于 FP-Growth 算法的关联性分析..... | 179 |
| 第 8 章 | 数据决策分析算法..... | 155 | 9.2.1 | 构建 FP 树..... | 179 |
| 8.1 | 基于 ID3 算法的决策分析..... | 156 | 9.2.2 | 频繁项分析..... | 181 |
| 8.1.1 | 信息量..... | 156 | 9.2.3 | 与 Apriori 算法比较..... | 184 |
| 8.1.2 | 信息熵..... | 156 | 9.3 | 基于 Eclat 算法的频繁项集挖掘..... | 184 |
| 8.1.3 | 信息增益..... | 157 | 9.4 | 本章小结..... | 185 |
| 8.1.4 | ID3 算法流程..... | 157 | 第 10 章 | 数据推荐算法..... | 187 |
| 8.1.5 | ID3 算法的应用..... | 157 | 10.1 | 概要..... | 187 |
| 8.2 | 基于 C4.5 算法的分类决策树 ... | 159 | 10.1.1 | 推荐算法发展..... | 188 |
| 8.2.1 | 概要..... | 159 | 10.1.2 | 协同过滤推荐..... | 189 |
| 8.2.2 | 应用实例..... | 159 | 10.2 | 基于 Item-Based 协同过滤推荐..... | 190 |
| 8.3 | 基于分类回归树的决策划分..... | 161 | 10.2.1 | Item-Based 基本思想..... | 190 |
| 8.3.1 | 概要..... | 162 | 10.2.2 | Slope One 实例：基于评分推荐..... | 190 |
| 8.3.2 | 应用实例：决策划分..... | 163 | 10.3 | 基于 User-Based 协同过滤推荐..... | 193 |
| 8.3.3 | 剪枝..... | 164 | 10.3.1 | 应用实例：根据人群的推荐..... | 194 |
| 8.4 | 基于随机森林的决策分类..... | 168 | 10.3.2 | User-Based 与 Item-Based 对比..... | 197 |
| 8.4.1 | 随机森林的特点..... | 169 | 10.4 | 基于潜在因子算法的推荐..... | 198 |
| 8.4.2 | 随机森林的构造方法..... | 169 | 10.4.1 | 应用实例：新闻推荐... .. | 198 |
| 8.4.3 | 应用实例：决定车厘子的售价层次..... | 170 | 10.4.2 | 流行度与推荐..... | 200 |
| 8.5 | 本章小结..... | 172 | 10.5 | 推荐算法与效果评价..... | 201 |
| 第 9 章 | 数据关联规则分析算法..... | 174 | 10.6 | 本章小结..... | 203 |
| 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 | | | |

第 1 章

算法基础

1.1 基础算法分析类型

算法是程序设计的灵魂，是数据处理基本思想体现。在算法的使用过程中，常常会涉及一些算法的分析思想，将这些思想进行归类整理后就可以形成算法的分析类型。常用算法分析类型包括分治法、动态规划法、回溯法、分支限界法及贪心法，它们都是解决不同问题的思想体现。

1.1.1 分治法

分治法是计算机领域中最常用的算法类型，核心思想在于“分而治之”，原理是首先将一个复杂混合的问题拆解为两个或者多个相同的子问题，然后将子问题依然进行分治法处理，直到子问题已经可以直接求解，最终按照问题的拆解顺序，从子问题开始逐步将问题合并为一个大的子问题，层层向上，最后得到该实际问题的解。在算法领域通常是越是规模小的问题越容易求解，分治法也是充分利用此优势发挥其价值。

利用分治法通常可以解决问题的类型有如下几种。

- (1) 待求解的问题是可以拆分为若干相同或者相似的子问题，子问题之间相互独立。

(2) 该问题可以在小规模情况下成立，并得到解。

(3) 问题被拆分后的子问题的解之间可以进行合并。

采用分治法解决问题的一般步骤如下。

(1) 分解问题：将原问题拆解为若干相同或相似、相互独立的子问题。

(2) 解决子问题：如果拆分的子问题依然无法直接获得解，则继续拆解子问题直到问题可以直接进行求解。

(3) 合并解：从可以直接求解的子问题入手，开始合并子问题的解，最终合并得到对原始问题的解。

分治法常常采用递归的方式进行求解，原因是递归可以有效地将问题拆解为相似且更小的子问题，并且子问题之间相互独立、互不干扰。例如，对二分查找、归并排序、快速排序等。但是，若利用分治法时，子问题之间存在重复子问题，则可能影响分治法的计算效率，这是因为相同的问题在不同的子问题中均被求解。在这样的情况下则可以考虑利用动态规划的方式对问题进行求解。

1.1.2 动态规划法

动态规划是一种将原问题拆解为若干相对简单子问题的求解方法，常常用于重叠子问题和最优子结构性质的问题。通过动态规划的方法，计算量则远远小于一般的解法。原因在于对于重叠子问题，一般情况下会被重复计算，而动态规划则是将重复计算简化为计算一次就放入到结果表中，在下一次计算时则从结果表中查询，从而直接获得结果，因此使性能得到提升。

动态规划的方法是一个从初始状态开始计算结果，后续的计算都依赖于前一个计算结果的状态，最终获得解的过程。主要步骤包括如下。

(1) 划分问题：根据问题的特征，将问题拆解为若干步骤，这些步骤中每个步骤应当清晰。

(2) 确定状态和状态变量：确定计算过程中需要记录的中间状态，并设置状态变量记

录中间状态。

(3) 确定状态转移过程：状态从初始状态开始，确定在计算过程中状态的转移方式。

(4) 寻找终止条件：动态规划不是一个无穷的过程，当状态转移到某个阶段时，触发终止条件得到问题最终的解。

但是并不是针对所有的重叠子问题都是有效的，它适用的情况包括如下情况。

(1) 最优子结构性质：最优子结构性质是一种最优化原理，表示如果一个问题的最优解所包含的所有问题的解也是最优的。

(2) 子问题重叠性质：子问题重叠表示在把一个大的问题拆解为若干子问题的过程中，在不同的子问题中会重复计算某些问题。动态规划的方法针对子问题重叠计算的问题，将每个子问题求解的结果存入子问题结果表中，当再次进行子问题计算时，首先从结果表中查询是否已经计算过，如果已经计算过则直接获取结果，如果没有则进行计算，并将计算的结果存入结果表中。

利用动态规划的方法可以解决背包问题、求解最长公共子序列、维特比算法等。

1.1.3 回溯法

回溯法是一种通过不断尝试获得问题解的方法。“回溯”的含义是当发现当前已经不满足求解的条件时，则采用“回溯”的方式，尝试其他的方法对问题求解，这种尝试的过程也是一种不断枚举的过程。

回溯法进行尝试的过程并不是漫无目的的穷举解的过程，它在选择求解目标过程中会利用当前条件与目标的最优匹配选择进行尝试，即使当前尝试失败，也可以回溯到上一步继续选择最后匹配选择进行。回溯法解体的过程相对比较通用，同生活中解决一般问题的情况基本一致，因此回溯法的应用相对较为广泛。

回溯法在所有的解空间中，按照深度优先尝试的原则，从初始状态不断深入到各个解空间中，解决过程如下。

(1) 对需要解决的问题，确定问题的解空间，确保在解空间的范围内存在最优解。

(2) 确定回溯法的尝试规则及路径寻找的方式，以减少不必要的计算。

1.1.4 分支限界法

分支限界法和回溯法是类似的问题求解方法。回溯法是通过深度优先的方式对问题进行探索性的解决，而分支限界法则是以广度优先或者以最小代价优先的方式尝试性解决问题。分支限界法的搜索策略是首先通过在路径选择处，生成其所有的下一步分支，计算每个分支的某个条件限定值，这些条件限定值即为分支的界限，然后从这些分支中选择某个的分支作为下一个节点，使问题的搜索路径向着最优解的分支进行，此处选择分支的方式主要有两种。

(1) FIFO (First In First Out): 按照先进先出的方式进行选择。

(2) 最小代价: 选择当前分支中代价最小的分支进行选择。这是一种基于优先队列的方式，优先选择某些分支进行搜索。

分支限界法的核心思想是：需要确定某个分支的上下界，一边搜索一边移除某些不满足条件的分支，以提升问题解决的效率。

用一个例子区分分支限界法与回溯法，即一个广度优先，一个深度优先。例如，将一堆质量相等的货物放到两辆货车中，若要确定最优的装载方案，则首先尽可能使第一辆货车装满，然后将剩余的货物装载到第二辆货车中，此种方法属于回溯法，该方法尽可能寻找与目标相近的解，因此会不断地往同一辆车中装载货物直到装满，然后换另外一辆货车。而对于分支限界法，则会选择每次最小代价，这种代价可以定义为每辆车剩余的装载量，所以会在两辆货车之间交替添加货物。

1.1.5 贪心法

贪心法也是比较常用的解决问题的方法之一。所谓“贪心”，是指当前求解的过程在当前的条件下是最好的选择，并未从整体上触发考虑，属于局部最优解。该方法没有一定固定的解题框架，重点在于贪心法的贪心策略。

贪心算法并不是一定会使所有问题得到全局最优的解，因此贪心算法适用于通过局部最优解可以获得全局最优解的问题。如果不满足此条件，则计算的结果可能与预期存在差