

“十三五”

国家重点图书

大数据科学丛书

从计算理论到 大数据图灵机

朱定局 著

高等教育出版社

“十三五”国家重点图书
大数据科学丛书

从计算理论到 大数据图灵机

朱定局 著

高等教育出版社·北京

图书在版编目(CIP)数据

从计算理论到大数据图灵机 / 朱定局著. — 北京 :
高等教育出版社, 2019.4

ISBN 978-7-04-051305-9

I. ①从… II. ①朱… III. ①数据处理 IV.
①TP274

中国版本图书馆 CIP 数据核字(2019)第 023941 号

策划编辑 冯 英
插图绘制 于 博

责任编辑 冯 英
责任校对 刘 莉

封面设计 张志奇
责任印制 陈伟光

版式设计 徐艳妮

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
印 刷 北京新华印刷有限公司
开 本 787mm×1092mm 1/16
印 张 14.5
字 数 260 千字
购书热线 010-58581118
咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.hepmall.com.cn>
<http://www.hepmall.com>
<http://www.hepmall.cn>

版 次 2019 年 4 月第 1 版
印 次 2019 年 4 月第 1 次印刷
定 价 49.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换
版权所有 侵权必究
物 料 号 51305-00

一直以来,计算理论由于理论性强,距离实际应用较远,而未受重视。随着大数据的发展,仅通过算法优化已很难解决大数据计算过程中的系列问题,而计算理论为解决这些问题提供了思路 and 方向,促进大数据技术走得更稳更远更好。

在大数据时代来临之前,人们对计算理论重视程度不够,主要是因为数据量不大时,数据是否可计算、是否是难解问题、计算复杂度是否高等因素,相对已有的计算资源影响不大。当数据量不大时,即使不能看出是否是可计算的问题,也能够通过归化到已有的不停机问题而推导出来;是否是 NP 问题或 NPC 问题,能够通过归化到已有的 NPC 或 NP 问题而推导出来;计算复杂度的高低也无关紧要,因为只要充分地进行算法优化,复杂度高的问题也能在比较短的时间内完成。而当数据量非常非常大时,计算复杂度将无法逆转,因为计算复杂度是一个趋势,而不是一个定量的值,当数据量不大时,这个趋势可以通过算法的优化来缩短计算时间,当数据量超级大时,将无法阻挡住这个大趋势。综上可见,计算理论对大数据的发展具有重要意义。

本书采用辩证发展科研创新法研究和讨论计算理论和大数据图灵机。其中,第 1—5 章是计算理论的部分,属于辩证发展科研创新法的思辨创新性阶段;第 6 章大数据应用部分是辩证发展科研创新法的实证创新阶段,是在大数据应用中对计算理论进行实证性创新;第 7 章大数据图灵机部分是辩证发展科研创新法的统一阶段,是前面两个阶段的综合性创新,使得创新从量变到质变,从继承性创新进展到了原始性创新,大数据图灵机就是通过辩证发展科研创新法实现的原始性创新。

本书是一本原创性的著作,通过苏格拉底、柏拉图式的问答法,引导读者得到知识,而不是直接把知识灌输给读者,使得读者不再是存储知识的“硬盘”,而是加工处理知识的“中央处理器”。这种科研创新方法使得学习者所掌握的知识不再是死的知识,而是可以在学习者大脑中加工处理的活知识。开篇从读者容易理解错的观点“计算理论就是关于计算方法的理论”切入,逐步引导读者到达“计算理论是关于计算能力的理论”的正确认知。本书循序渐进、环环相套,逻辑性强,对计算理论和大数据一无所知的读者,从头读起,也能渐入佳境,从而了解计算理论和大数据的知识。

本书介绍的研究成果得到了国家社会科学基金重大项目(14ZDB101)、国家自然科学基金重点项目(41630635)、教育部-腾讯公司产学研合作协同育人项目(201602001001)、国家级新工科研究与实践项目(教高厅函[2018]17号)、广东高校重大科研项目(2017KTSCX048)、广东省新工科研究与实践项目(粤教高函[2017]118号)、广东省高等教育教学研究和改革重点项目(粤教高函[2016]236号)、广东省学位与研究生教育改革研究重点项目(粤教研函[2016]39号)、广东省联合培养研究生示范基地(粤教研函[2016]39号)、广东省产学研合作重大项目(2014B090901064)、华南师范大学校级质量工程建设项目规划教材项目(教学[2016]78号)、华南师范大学校级质量工程建设项目教学团队项目(教学[2016]78号)、华南师范大学校级质量工程建设项目精品教材项目(教学[2014]52号)的支持。由于作者水平有限,书中难免有不妥甚至错误之处,恳请读者批评指正。

朱定局

2019年1月于华南师范大学

目录

第 1 章	计算理论概述	1
1.1	什么是计算理论	1
1.2	计算理论研究什么	3
1.3	为什么需要计算理论	4
1.4	怎么学习计算理论	5
第 2 章	计算模型	7
2.1	图灵是谁	7
2.2	中国的“图灵”	8
2.3	从算盘中抽象图灵机	11
2.4	形式化地表示图灵机	13
2.5	图灵机的运行	18
2.6	图灵机的格局	20
2.7	图灵机定义与每一次计算之间的关系	23
2.8	设计图灵机实体	25
2.9	怎样解决图灵机运行结果的错误	31
2.10	图灵机如何支持各种程序的运行	37
2.11	通用图灵机和专用图灵机的实现	40
2.12	图灵机与图灵机实现之间的对应关系	42
2.13	图灵机的运行	46
第 3 章	可计算性与计算复杂性概述	49
3.1	不同图灵机实现的计算能力有没有差异	49
3.2	如何衡量计算能力	51
3.3	为什么所有计算机的计算能力与图灵机相同	52
第 4 章	可计算性	61
4.1	图灵机是否会不停机	61
4.2	停机问题的解决	62
4.3	什么是可计算性	66
4.4	判定问题与计算问题的关系	70

4.5	一个问题是否可判定的证明	71
4.6	与停机问题等价的问题	74
4.7	可判定、半可判定、不可判定之间的关系	80
4.8	可判定、不可判定是否对应可计算、不可计算	84
4.9	可计算理论的意义	85
第5章	计算复杂性	86
5.1	什么是计算复杂性	86
5.2	为什么判定性问题的算法复杂性对计算性问题 同样适用	87
5.3	如何衡量复杂程度	89
5.4	非确定型图灵机和确定型图灵机的区别	91
5.5	在多项式时间内猜出 NP 问题的解	92
5.6	如何不猜解求解 NP 问题	94
5.7	非确定型图灵机与确定型图灵机是否等价	95
5.8	P 问题和 NP 问题的关系	97
5.9	P 问题和 NP 问题对应的计算性问题	98
5.10	把可计算问题划分成 P、指数型、NP、NPC、 NPH 问题	99
5.11	证明一个问题是 NPC 问题	100
5.12	定量地表示算法的时间复杂度	103
5.13	常见的算法时间复杂度	104
5.14	非多项式时间复杂度与多项式时间复杂度	110
5.15	多项式、非多项式时间复杂度与 P、NP 问题的 关系	111
5.16	不同时间复杂度的比较	112
5.17	复杂度的形式化表示	113
5.18	算法复杂度的本质	114
5.19	时间复杂度和空间复杂度	116
5.20	关系复杂度	116
5.21	复杂算法的分解	117
5.22	确定问题的规模 n	119
5.23	在规模相等的情况下的非多项式时间和多项式时间	119
5.24	降低算法的复杂度	120

5.25	问题复杂度和算法复杂度的区分	122
第 6 章	图灵机的大数据应用	124
6.1	大数据的特性	124
6.2	大数据应用对图灵机的需求	127
6.3	大数据应用图灵机	129
6.4	跳板大数据应用图灵机	130
6.5	耦合大数据应用图灵机	134
6.6	先验大数据应用图灵机	139
6.7	自适应大数据应用图灵机	151
6.8	增量大数据应用图灵机	154
6.9	自动大数据应用图灵机	158
6.10	分治大数据应用图灵机	165
6.11	冗余大数据应用图灵机	168
第 7 章	大数据图灵机	173
7.1	大数据图灵机的基本模型	173
7.2	大数据图灵机的可计算性	180
7.3	大数据的计算复杂度、存储复杂度和数据复杂度	182
7.4	数据计算复杂度的大 O 表示法	188
附录 1	辩证发展科研创新法	191
附录 2	习题	197
附录 3	模拟题库	203
参考文献		216

第 1 章 计算理论概述

计算理论在很多领域都有着广泛的应用^[1-11],但这些计算理论和本书中所讲的计算理论不同之处在于是否结合了具体应用。面向应用的计算理论需要结合应用的场景和需求,而本书中所讲的计算理论是计算机的计算理论,是抽象的、通用的计算理论,主要由计算模型、可计算性和计算复杂性三个部分组成。应用领域的计算理论离不开计算机的计算理论支撑,计算机的计算理论是根本和核心。面向应用的计算理论往往比较具体且有实际应用效果,所以研究的人比较多,而计算机的计算理论虽然也有研究^[12-14],但相对较少。计算机的计算理论在第一台电子计算机出现之前就已经被研究,相对比较成熟,图灵关于图灵机的论文论述就是关于计算机的计算理论,在计算机的计算理论方面有着很多经典著作^[15-20]。但随着计算机的发展,计算机的计算理论也有待进一步发展。本书针对计算机在大数据领域中的发展,对计算机的计算理论进行研究,探索计算机的计算理论如何更好地支撑大数据技术。

1.1 什么是计算理论

【老师】:

什么是计算理论?

【学生】:

计算方法有九九乘法口诀表、数学加法口诀、手指算进位加口诀等,最简单的计算如 $3+8=11$ 。计算理论是研究人类面对 $3+8$ 的问题如何得到 11 的计算方法。

【老师】:

错! 计算理论不是研究人类的计算方法。

【学生】:

难道计算理论是研究非人类的计算方法? 除了人类,机器人也可以进行计算,而机器人进行计算是依靠机器人的大脑——计算机,那么计算理论是研究计算机的计算方法?

【老师】:

错! 计算理论也不是研究计算机的计算方法。

【学生】:

计算理论既不是研究人类的计算方法,又不是研究非人的计算方法,那是研究什么的呢?

【老师】:

计算理论研究的不是计算方法。

【学生】:

那么计算理论研究的是计算的公式,还是计算的算法? 或者是计算的程序,再或者是计算的软件?

【老师】:

计算理论研究的不是计算的公式,也不是计算的算法、计算的程序和计算的软件,因为这些都是计算方法。

答案在哪里? 可以从人类自身寻找答案。当人类掌握乘法口诀之后就能够进行乘法计算,设想如果是狗掌握了乘法口诀后,能进行乘法计算吗? 结果是不能,是为什么哪?

【学生】:

可能的原因是狗记不住乘法口诀。

【老师】:

那鹦鹉能不能记得住呢? 人们常见鹦鹉学舌,能背出乘法口诀,那么鹦鹉能进行乘法计算吗?

【学生】:

不能。

人和鹦鹉虽然都能记住计算方法的口诀,但前者能进行计算,后者却不能进行自主计算,区别在哪里?

【老师】:

区别在人和鹦鹉本身的不同。因为计算与大脑有关,所以问题的关键在于人的大脑和鹦鹉的大脑的不同。

【学生】:

不同在哪里?

【老师】:

鹦鹉的大脑不具备对计算方法的理解和处理能力;而人的大脑具备这种能力,这种能力就是计算能力。计算方法只是计算的知识。大家都知道能力比知

识重要,同样计算能力也比计算知识重要。正如成年人具备计算能力,也具备计算的知识,所以就能够通晓计算;婴儿具备计算能力但不具备计算的知识,虽然暂时不会计算,但长大以后经过学习就能掌握计算;而鹦鹉不具备计算能力,即使能记住计算的知识,也永远不会计算。将计算知识输入存储到大脑容易,要改变大脑的计算能力就难了,因为大脑的计算能力是由大脑本身的结构决定的。人类的大脑是在进化中逐渐形成了计算能力的,而鹦鹉的大脑还没有进化出计算能力。

【学生】:

看来计算能力的确比计算的知识更重要。那么,计算理论研究的是人脑的计算能力吗?

【老师】:

错。计算机专业研究的计算理论是计算机的计算能力。

我们来做一个游戏。

【第1步】

请一些同学分别扮演人的计算方法、计算机的计算方法、算法、程序、公式和计算机的计算能力。

请一位同学上来扮演计算理论。

【第2步】

请扮演计算理论的同学询问每个代表的名称,每个代表要回答询问,然后计算理论代表挑出与自己匹配的代表。

1.2 计算理论研究什么

【学生】:

计算理论是怎么研究计算机的计算能力的?

【老师】:

计算理论从三个方面研究计算机的计算能力:

① 如何使计算机具备计算能力——使计算机能进行计算的模型,简称计算模型。

② 计算机计算能力的适用范围——哪些是能计算的、哪些是不能计算的,简称可计算性。

③ 计算机计算能力的需求分析——要用多少时间、要用多少存储空间,简称计算复杂性。

我们来进行一个游戏。

【第 1 步】

请 4 位同学分别扮演计算理论、计算模型、可计算性和计算复杂性。

【第 2 步】

提出不同问题,请不同代表回答是否属于自己回答的范畴。

【第 3 步】

提出问题:怎么才能具备计算能力?

请计算理论代表回答。

请计算模型代表回答。

【第 4 步】

提出问题:人的计算能力是无限的吗?

请计算理论代表回答。

请可计算性代表说我能回答。

【第 5 步】

提出问题:试算 10000 之内有多少质数需要多长时间?

请计算理论代表回答。

请计算复杂性代表回答。

【第 6 步】

提出问题:计算机是怎么制造出来的?

请计算理论代表回答。

请计算模型代表回答。

【第 7 步】

提出问题:有没有计算不了的问题?

请计算理论代表回答。

请可计算性代表回答。

1.3 为什么需要计算理论

【学生】:

计算理论是关于计算机计算能力的理论,为了研究计算机的计算能力,所以需要计算理论进行研究。

【老师】:

错。为什么需要计算理论,不是为了研究计算机的计算能力。虽然计算机的计算能力是计算理论的研究内容,但“为什么需要计算理论”的实质是“计算机为什么需要计算能力?”

【学生】:

如果计算机没有计算能力,那就称不上计算机了。

【老师】:

既然计算机没有计算能力就称不上是计算机,那么如果没有计算理论,还会有计算机么?

【学生】:

没有。

【老师】:

对。没有计算理论就不会有计算机。正是因为我们需要计算机,所以我们需要计算理论。现有的计算机都是以计算理论为基础研制出来的。

1.4 怎么学习计算理论

【学生】:

是不是要通过计算机的构造来学习计算理论呢?

【老师】:

这涉及怎么学习计算理论的问题。首先,计算机类型有很多,如台式机、笔记本式计算机、超级计算机。如果只学会如何构造台式机,还是不会构造笔记本式计算机和超级计算机,所以只能逐一类型去学习。其次,电子计算机从诞生到现在也有成千上万种。

【学生】:

能不能只学习其中一种类型的计算机?

【老师】:

不能。因为不同类型的计算机之间有差异,一种类型的计算机代表不了其他类型的计算机。

【学生】:

有没有一种可以代表所有类型的计算机?

【老师】:

这种计算机过去没有,现在、将来也没有。世界上两片相同的树叶都没有,

更何况是比树叶复杂得多的计算机,所以不可能有代表所有其他类型的计算机。但是,可以把所有类型计算机的本质抽象出来,形成一个计算模型。这样,就可以通过学习计算模型的构造,了解电子计算机的构造。

第2章 计算模型

计算模型在很多领域都有应用^[21-34],而对计算机的计算模型的研究相对较少^[35-43]的原因,是因为计算机的计算模型(例如图灵机)已经比较成熟。虽然计算机技术日新月异,但图灵机仍然适用,这是计算机的计算模型优势所在,不依赖于具体的硬件和软件,具有较高的抽象性,能刻画计算机的本质,即使计算机从超级计算机发展到嵌入式计算机,仍然没有超过计算机的计算模型图灵机的刻画范围。对计算机计算模型的研究主要是对图灵机的研究,图灵机是最经典的计算模型,也是从直观上最容易理解和最简洁的计算模型,是由图灵(Alan Mathison Turing)所提出来的一种计算模型。

2.1 图灵是谁

【学生】:

只需要学习一个计算模型的构造就等于学习了成千上万种计算机的构造,实在是太好了。老师真聪明!

【老师】:

不是我聪明,是图灵聪明。

【学生】:

图灵是谁?

【老师】:

图灵是计算理论的先驱。他在创建计算理论的时候,并没有通过构造一台计算机来表示计算理论,而是通过构造一个计算模型来表示计算理论。因此,图灵被称之为计算机科学之父、计算机逻辑奠基者。

【学生】:

图灵为什么被称为计算机科学之父?图灵没有构造出计算机实体,只构造了计算模型。

【老师】:

虽然计算机的主板、中央处理器、存储器(内存、硬盘)等千变万化,但计算

模型都是一样的。而计算机的这个计算模型是图灵构造出来的,所以图灵被称为计算机科学之父。

【学生】:

好崇拜图灵,真想去拜访他。

【老师】:

图灵出生于 1912 年,已经一百多岁了。他已经不在人间,不过可以通过照片认识他。

【学生】:

从照片看图灵不像是中国人?

【老师】:

图灵是英国人。不过中国也有一个“图灵”。



图 2.1 图灵

2.2 中国的“图灵”

【学生】:

中国的“图灵”是谁?

【老师】:

中国的“图灵”在约 1800 年前制造出了“世界上最古老的计算机”,他就是刘洪。

【学生】:

可是现在的计算机都是根据图灵的计算模型制造出来的,1800 年前图灵还没有出生,又怎么会有计算机?

【老师】:

并不矛盾。中国古代制造的“计算机”是手动的,现在的电子计算机是自动的。

【学生】:

既然中国古代也制造出了“计算机”,那么我们为什么不学习中国“图灵”的计算模型?

【老师】:

中国“图灵”虽然制作出了“世界上最古老的计算机”,但并没有提出计算模型,所以在中国也就没有发展出千变万化的现代计算机。

可见,提出一个计算模型比制作出一台计算机的意义更为重大。

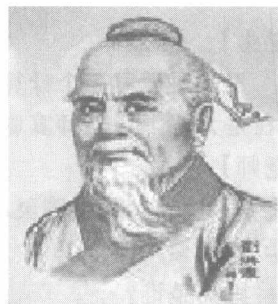


图 2.2 刘洪

【学生】:

既然中国“图灵”没有提出计算模型,为什么在没有计算模型的情况下也能制造出“计算机”呢?

【老师】:

中国“图灵”在制作“计算机”的时候,心中应该已经有了一个计算模型,只不过他没有把这种计算模型表达和抽象出来,否则我们现在学习的就是中国“图灵”的计算模型了。

【学生】:

虽然计算模型不是中国人提出来的,但中国人能制作出“最古老的计算机”也是非常自豪的。这台古老的“计算机”是不是只有在博物馆里才能看到?

【老师】:

这种古老的“计算机”不但在博物馆里能看到,现在仍有人在使用,而且你们上小学的时候还使用过。

【学生】:

已经 1800 年了,现在还有人在使用,这种“计算机”的生命周期也太长了。现代的电子计算机更新换代速度非常快,这种古老的“计算机”是什么样子?我想看看。

【老师】:

就是这个样子的。

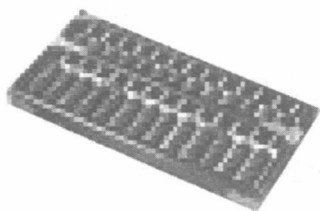


图 2.3 算盘

【学生】:

这不是算盘吗? 它和计算机一点都不像,为什么说是计算机呢?

【老师】:

判断是否是计算机的标准不在于它的外形。例如,笔记本式计算机的外形和台式计算机的外形和超级计算机的外形也是天壤之别,但它们全部都是计算机。是否是计算机的唯一判断标准,在于它是否满足图灵提出的计算模型。准