



智能

科/学/技/术/著/作/丛/书

粒计算与数据推理

闫林 闫硕 著



科学出版社

智能科学技术著作丛书

粒计算与数据推理

闫林 闫硕 著



科学出版社

北京

内 容 简 介

本书内容围绕数据推理展开,旨在采用推理的形式描述数据联系,以支撑问题程序化的算法设计。运作于数据之间是各推理方法的推演特点,是区别于各逻辑推理系统的主要标志。为使不同形式的数据推理融为整体,本书贯穿了粒计算的数据处理理念。第1章针对粒计算,设立了粒的形式化框架,统一了各种粒的定义,明确了粒计算的讨论方式。在此基础上,第2~7章建立了体现粒计算内涵的数据推理方法,主要包括粗糙数据推理的定义和性质分析、粗糙数据推理蕴含的粗糙数据联系、粗糙数据推理的度量讨论、决策推理与决策系统化简、数据关联推理与数值信息、数据合并的结构表示和粒化处理、结构转换的矩阵变换和代数计算、基于关系结构的三支决策、实际问题描述等。各种方法与数据的关联、蕴含、组合或转换等数据推理问题相关,形成了贯穿粒计算理念的数据推理的内容体系。

本书可作为计算机专业的科研人员、高校师生的参考书,也可供从事粒计算、数据处理、数据挖掘、机器学习等人工智能领域研究的学者参考。

图书在版编目(CIP)数据

粒计算与数据推理 / 闫林, 闫硕著. — 北京: 科学出版社, 2019.6

ISBN 978-7-03-061259-5

(智能科学技术著作丛书)

I. ①粒… II. ①闫… ②闫… III. ①人工智能—计算方法
IV. ①TP18

中国版本图书馆CIP数据核字(2019)第094699号

责任编辑: 张海娜 纪四稳 / 责任校对: 严 娜

责任印制: 师艳茹 / 封面设计: 陈 敬

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

天津文林印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2019年6月第 一 版 开本: 720 × 1000 1/16

2019年6月第一次印刷 印张: 17 3/4

字数: 350 000

定价: 98.00 元

(如有印装质量问题, 我社负责调换)

《智能科学技术著作丛书》编委会

名誉主编：吴文俊

主 编：涂序彦

副 主 编：钟义信 史忠植 何华灿 何新贵 李德毅 蔡自兴 孙增圻
谭 民 韩力群 黄河燕

秘 书 长：黄河燕

编 委：(按姓氏汉语拼音排序)

蔡庆生(中国科学技术大学)

蔡自兴(中南大学)

杜军平(北京邮电大学)

韩力群(北京工商大学)

何华灿(西北工业大学)

何 清(中国科学院计算技术研究所)

何新贵(北京大学)

黄河燕(北京理工大学)

黄心汉(华中科技大学)

焦李成(西安电子科技大学)

李德毅(中央军委联合参谋部第六十一研究所)

李祖枢(重庆大学)

刘 宏(北京大学)

刘 清(南昌大学)

秦世引(北京航空航天大学)

邱玉辉(西南大学)

阮秋琦(北京交通大学)

史忠植(中国科学院计算技术研究所)

孙增圻(清华大学)

谭 民(中国科学院自动化研究所)

谭铁牛(中国科学院自动化研究所)

涂序彦(北京科技大学)

王国胤(重庆邮电学院)

王家钦(清华大学)

王万森(首都师范大学)

吴文俊(中国科学院数学与系统科学研究院)

杨义先(北京邮电大学)

于洪珍(中国矿业大学)

张琴珠(华东师范大学)

赵沁平(北京航空航天大学)

钟义信(北京邮电大学)

庄越挺(浙江大学)

《智能科学技术著作丛书》序

“智能”是“信息”的精彩结晶，“智能科学技术”是“信息科学技术”的辉煌篇章，“智能化”是“信息化”发展的新动向、新阶段。

“智能科学技术”(intelligence science & technology, IST)是关于“广义智能”的理论方法和应用技术的综合性科学技术领域，其研究对象包括：

- “自然智能”(natural intelligence, NI)，包括“人的智能”(human intelligence, HI)及其他“生物智能”(biological intelligence, BI)。

- “人工智能”(artificial intelligence, AI)，包括“机器智能”(machine intelligence, MI)与“智能机器”(intelligent machine, IM)。

- “集成智能”(integrated intelligence, II)，即“人的智能”与“机器智能”人机互补的集成智能。

- “协同智能”(cooperative intelligence, CI)，指“个体智能”相互协调共生的群体协同智能。

- “分布智能”(distributed intelligence, DI)，如广域信息网、分散大系统的分布式智能。

“人工智能”学科自 1956 年诞生以来，在起伏、曲折的科学征途上不断前进、发展，从狭义人工智能走向广义人工智能，从个体人工智能到群体人工智能，从集中式人工智能到分布式人工智能，在理论方法研究和应用技术开发方面都取得了重大进展。如果说当年“人工智能”学科的诞生是生物科学技术与信息科学技术、系统科学技术的一次成功的结合，那么可以认为，现在“智能科学技术”领域的兴起是在信息化、网络化时代又一次新的多学科交融。

1981 年，中国人工智能学会(Chinese Association for Artificial Intelligence, CAAI)正式成立，25 年来，从艰苦创业到成长壮大，从学习跟踪到自主研发，团结我国广大学者，在“人工智能”的研究开发及应用方面取得了显著的进展，促进了“智能科学技术”的发展。在华夏文化与东方哲学影响下，我国智能科学技术的研究、开发及应用，在学术思想与科学方法上，具有综合性、整体性、协调性的特色，在理论方法研究与应用技术开发方面，取得了具有创新性、开拓性的成果。“智能化”已成为当前新技术、新产品的发展方向和显著标志。

为了适时总结、交流、宣传我国学者在“智能科学技术”领域的研究开发及应用成果，中国人工智能学会与科学出版社合作编辑出版《智能科学技术著作丛

书》。需要强调的是，这套丛书将优先出版那些有助于将科学技术转化为生产力以及对社会和国民经济建设有重大作用和应用前景的著作。

我们相信，有广大智能科学技术工作者的积极参与和大力支持，以及编委们的共同努力，《智能科学技术著作丛书》将为繁荣我国智能科学技术事业、增强自主创新能力、建设创新型国家做出应有的贡献。

祝《智能科学技术著作丛书》出版，特赋贺诗一首：

**智能科技领域广
人机集成智能强
群体智能协同好
智能创新更辉煌**

涂序彦

中国人工智能学会荣誉理事长

2005年12月18日

前 言

近年来,人工智能成果可喜,影响程度前所未有。与人工智能相关的产品、服务或技术已渗透到社会的各个方面,使与智能处理联系在一起的信息产业、现代农业、工业制造、服务行业等得到迅猛的发展。

人工智能的名称包含了智能控制或自动处理的内涵,对其研究和开发的目的也是如此。要想达到目标,相关研究方法的构建是任何智能处理必须面对的任务。由于不同的问题涉及不同的策略,例如,机械操控与围棋对弈显然依托不同的方法,加之各类问题纷繁复杂,故人工智能是涵盖众多科研方向的交叉学科领域。在人工智能涉及的众多领域中,信息科学与人工智能密切相关,被视为人工智能领域的核心内容。而计算机科学是信息科学的重要组成部分。

本书针对计算机科学的问题,围绕描述方法展开讨论,与智能处理的基础理论关联在一起。计算机科学包含众多的分支和方向,无论它们针对的问题如何,其目标方向是一致的,即依据一定的方法使问题得到程序化的处理。本书的讨论旨在建立数据处理的方法,主要面对数据之间的组合、关联、转换或蕴含等数据联系问题,并把存在差异的各种数据联系统一称为数据推理。粒计算是许多研究者关注的课题,其核心内涵体现在从数据整体中获取一类构成粒的数据,再通过粒的操作分析完成数据的处理。本书将粒计算的数据处理方式贯穿始末,使各种数据推理方法在其内涵下得到统一。

本书共7章。第1章介绍粒计算的概念,其重点集中于粒的定义,同时对粒的组合运算给予概括性的分析,确立针对数据推理问题撰写的依据。第2~7章围绕数据推理这个主题,把推理的方法与粒的组合运算联系在一起,形成以粒计算为基础、以数据推理为主题的内容体系。

本书通过理论研究,提出数据处理方法,并将其作为算法设计的支撑,以求问题的程序化处理。本书的几种数据推理方法融入粒计算数据处理的思想,可看成粒计算数据处理的途径。尽管本书涉及理论研究,并以推理的形式出现,但是各种数据推演均源于实际中的数据关联现象,这些现象都是数据处理面对的问题。因此,将提出的方法用于实际问题的描述,使理论应用于实际也是撰写本书的初衷。

限于作者水平,书中难免存在疏漏之处,谨与读者交流,希望不吝指正。

目 录

《智能科学技术著作丛书》序

前言

第1章 粒与粒计算的概念	1
1.1 数据与数据推理概述	1
1.2 粒和粒计算的直观解释	4
1.2.1 直观意义下的粒	5
1.2.2 直观意义下粒之间的计算或运算	7
1.2.3 直观意义下的粒度问题	10
1.3 粒的形式化讨论	11
1.3.1 数据集	12
1.3.2 粒的形式化方法分析	14
1.3.3 粒的形式化框架及分析	15
1.4 粒的几种形式化方法	18
1.4.1 相关的符号和概念	18
1.4.2 粗糙集方法确定的粒	21
1.4.3 信息系统确定的粒	24
1.4.4 关系确定的粒	27
1.4.5 融合信息确定的粒	28
1.4.6 划分确定的粒	29
1.5 等价关系和划分	32
1.6 粒的进一步讨论	34
1.6.1 基于数据集 $U \times U$ 上的粒	35
1.6.2 决策系统的分解与粒的产生	36
1.7 粒的计算或运算和粒度问题	39
1.7.1 粒之间计算或运算的相关问题	39
1.7.2 粒度和粒度变化问题	41
第2章 粗糙数据推理	44
2.1 粗糙推理空间和粗糙数据推理的定义	45
2.1.1 粗糙数据联系	45
2.1.2 粗糙推理空间的构建	47
2.1.3 粗糙数据推理的定义	49

2.2	关于 R 的粗糙数据推理	53
2.2.1	准备工作及相关概念	53
2.2.2	粗糙数据推理的性质	55
2.3	粗糙关系与粗糙数据推理	58
2.3.1	粗糙关系	58
2.3.2	基于粗糙关系的粗糙数据推理	59
2.4	不同的等价关系与粗糙数据推理	62
2.4.1	基本性质	63
2.4.2	基于不同等价关系的粗糙数据推理	64
2.4.3	嵌入算法与粗糙数据推理的判定	65
2.5	粗糙数据推理的应用	67
2.5.1	实际问题讨论	67
2.5.2	进一步的说明	69
2.6	树型推理空间与粗糙数据推理	71
2.6.1	树型推理空间	72
2.6.2	分层推理空间	74
2.6.3	分层推理空间中的粗糙数据推理	75
2.6.4	分层推理空间的扩展及粗糙数据推理	76
2.7	分层推理空间的细化	80
2.7.1	细化分层空间	80
2.7.2	细化分层空间中的粗糙数据推理	83
2.8	实际问题描述	85
2.8.1	分层推理空间和细化分层空间的构建实例	85
2.8.2	实际问题讨论	91
第3章	粗糙数据推理的度量	101
3.1	粗糙数据推理的内涵问题	101
3.2	粗糙路径	103
3.2.1	粗糙路径的定义	103
3.2.2	粗糙路径的解释	105
3.3	粗糙路径的性质	109
3.4	粗糙路径与粗糙数据推理	112
3.5	粗糙数据推理的精度	115
3.6	精度的性质	121
3.7	粗糙数据推理的连接度	126
3.7.1	连接度的定义	127
3.7.2	连接度的性质	129

3.7.3 度量与粒计算之间的联系	132
第 4 章 决策推理与决策系统的化简	134
4.1 基于决策推理的决策系统分解	135
4.1.1 决策系统的结构与数据的取值	135
4.1.2 公式和粒	138
4.1.3 决策公式和决策推理	140
4.1.4 决策系统的分解	143
4.1.5 决策确定与否的判定	146
4.2 样本的分解匹配处理	150
4.2.1 决策系统和样本	151
4.2.2 样本和匹配	152
4.2.3 样本和决策	155
4.2.4 具体的匹配讨论	159
4.3 基于决策推理的数据合并及其决策系统化简	161
4.3.1 属性约简概述	162
4.3.2 数据合并及其简化系统	165
4.3.3 属性子集及其合并方案集	168
4.3.4 A -集合及其简化式	170
4.3.5 决策系统的等价性讨论	172
4.3.6 简化式的进一步讨论	177
第 5 章 基于粒化树的数据关联	181
5.1 数据关联的直观解释	181
5.2 粒化树的构建	184
5.2.1 预备工作	184
5.2.2 粒化树构建的直观方法	185
5.2.3 粒化树构建的形式化方法	187
5.3 两个数据集之间的数据关联	191
5.3.1 数据关联的定义	191
5.3.2 数据关联的性质	192
5.3.3 数据关联之间的联系	194
5.4 实际问题描述	197
5.5 基于一个数据集的数据关联	202
5.5.1 划分与相关集类	202
5.5.2 相关集类与粒化树	205
5.5.3 相关集类与数据关联	206
5.5.4 数据关联的性质	207

5.5.5 实际数据关联的讨论	209
第 6 章 结构化的数据合并与矩阵变换	215
6.1 数据合并问题	215
6.2 结构转换与数据合并	218
6.3 矩阵表示和矩阵变换	222
6.3.1 关联结构和关联矩阵	223
6.3.2 粒化结构和粒化矩阵	225
6.3.3 矩阵的行相加处理	228
6.3.4 矩阵的列合并处理	230
6.4 实际问题描述	233
6.5 结论的证明	235
6.6 无关权的数据合并说明	242
第 7 章 基于关系结构的三支决策	245
7.1 三支决策的解释	245
7.2 三支划分	247
7.3 关系结构和三支划分	249
7.3.1 关系结构的构建	250
7.3.2 三支划分的构建	251
7.3.3 三支划分的性质	255
7.4 基于三支划分的决策结果	257
7.4.1 决策结果的层次信息	257
7.4.2 决策结果之间的联系	259
7.5 实际问题描述	262
7.6 三支决策方法的总结	265
参考文献	267
后记	270

第1章 粒与粒计算的概念

本书各章节的讨论将针对数据之间的推出、关联、组合或蕴含等涉及数据处理的一些问题，可归结如下研究内容：数据之间的非精确推理、非精确数据推理的度量描述、决策推理与决策系统的分解与化简、基于公共数据的数据关联、结构化的数据合并与矩阵计算、以数学结构为支撑的三支决策推理、实际问题的刻画描述等。这些内容都与数据之间的相互联系或数据的特性密切相关，既涉及信息科学分支中的问题，也源于实际中的数据关联现象。又因为各类数据联系以及针对数据特性的讨论可以广义地看作数据之间的推理，所以本书把涉及的内容统一称为数据推理，数据推理是把各种数据关联形式归为统一主题的名称。虽然数据推理将涉及数据之间不同的联系形式，或涉及数据特性的讨论分析，但各章节的内容将通过粒计算的数据处理方式联系在一起，粒计算的数据处理内涵将作为贯穿本书的主线。

于是我们自然提出这样的问题：什么是粒？粒计算是怎样形式的运算？实际上，虽然粒计算是近年来信息科学关注的研究热点，并从认识方面具有了一定的共识或得到直观意义上的认同，但从数学描述或严格定义方面考虑，粒计算仍需进行深入和系统的研究，以达到建立理论体系的目的。近年来，粒计算的研究取得了许多成果，已建立了粒计算研究的相关方法。因此，为了使各章节针对数据推理的讨论在粒计算数据处理内涵下联系在一起，本章先对数据推理涉及的内容进行概述，然后介绍粒计算的相关概念。

1.1 数据与数据推理概述

在信息技术与社会发展高度融合的今天，针对各类问题的程序化处理与问题涉及的对象密切相关。这里的对象既包括实际中客观存在的各类实物，又涉及各类实物的抽象化表示。例如，问题程序化过程中常常面对的人类、动物、物体、植物等具体存在的实物，以及为了描述这些实物采用的数字、字母、符号、像素等抽象的概念等，它们都是程序化过程不可避免的对象，并与我们的生活息息相关。实际上，对实物的形式化描述是对问题程序化的前提，形式化是指对实际事物的抽象化或符号化表示，只有把实际物体实施抽象化表示后，才可能通过数学方法进行描述，并通过编制程序予以处理。因此，实施程序化处理的前提是对各类对象的形式化或抽象化表示。例如，为了对一类学生（即一群人）进行自动化或

程序化的管理,我们往往把每一位学生对应一个学号,数字化的学号就是对具体对象(即学生)的抽象化表示。又如,对实际对象进行计算机识别时,需要记录对象的数字图像信息,数字图像信息就是实际对象的形式化或抽象化的表示,也称为符号化表示。

在下面的讨论中,我们把实际对象的抽象表示称为数据,即数据是实物的形式化的表示。但在讨论问题时,我们往往直接把具体的实物称为数据。例如,把学生、树木、工厂、站点等这些具体存在的对象都视为数据。严格上讲,只有把这些实际对象形式化或符号化后,符号化的表示才称为数据。不过直接把实物称为数据符合人们的习惯,不会产生麻烦,人们也常常如此看待数据。

数据是信息科学研究处理的对象,是程序化过程中各类操作的支撑,是信息科学中各类问题研究、处理、操作、存储的基础信息。

针对数据的各类研究或操作就是数据处理问题,并可给出一般性的概括:数据处理是指对数据之间的联系、数据之间的组合、数据之间的蕴含、数据之间的推理、数据之间的合并、数据信息的识别、数据构成的结构、数据的约简和存储、数据自身的性质等各类涉及数据的问题进行描述、刻画、分析或程序化的方法。因此,涉及对数据的讨论、操作、处理或分析等方面的工作都可以归为数据处理的范畴。

本书的讨论将围绕数据处理的相关问题,虽然涉及不同的方面,但可归结为统一的主题——数据推理。下面各章节的讨论将建立一些数据推理的方法,并体现自身的研究特点。如果直观地给予解释,那么数据推理可包括以下方面:数据之间的联系、数据之间的蕴含、数据之间的依赖、数据之间的关联、数据之间不明确的联系、数据之间的组合、数据的合并化简、数据的特性分析、数据引出的决策等,这些都与数据之间的联系密切相关。在某种意义上,数据联系可以广义地认为是数据之间的推理,称为数据推理。因此,数据推理是讨论研究数据之间明确或不明确联系、蕴含、组合、化简的课题。

后面各章节的讨论都将围绕数据推理问题展开,之所以将数据推理作为讨论的课题,既缘于信息科学包含的问题,又来自实际当中数据之间明确或不明确的数据联系,也与实际中的数据合并重组、数据的分类处理、数据自身的特性等问题相关。例如,如果公交站点1与公交站点2相连,且公交站点2与公交站点3相连,那么公交站点1与公交站点3通过公交站点2的相连关系是明确的。如果把公交站点看作数据,则站点之间的连接关系反映了明确的数据联系。又如,如果张三把钱借给李四,并且李四把钱借给王五,则从张三经李四到王五的借贷关系是明确的,而张三的儿子经李四到王五的借贷关系虽然不明确,但是从人们通常的认识方面考虑,这种儿子继承的借贷联系往往被认可,展示为不明确的借贷关系。因此,如果把张三、李四、王五以及张三的儿子看作数据,那么张三的儿

子与王五的借贷关系展示了不明确的数据联系。在实际中,明确或不明确的数据联系还存在其他的形式,下面的讨论将对某些形式的数据联系展开研究。如果再对实际问题进行观察,我们还可以看到数据联系的另一现象——数据的合并重组,如两个企业的合并、一些高校的重组、城市群的建设、某群人的归类等,这些都展示了数据联系的另一形式,也常在信息科学的研究中涉及。因此,对数据联系的模式进行刻画,建立描述方法,是算法设计或问题程序化的前提,对于实际问题的智能处理具有理论支撑和实际应用的意义。

同时我们注意到,很多情况下,在数据联系确定的关联之中,包含着联系的方向。例如,从张三经李四到王五明确的借贷关系或从张三的儿子经李四到王五不明确的借贷关系中包含从张三或其儿子到王五的借贷方向,同样公交车从站点到站点的行驶也离不开方向问题。这启发了利用推理方法刻画数据联系的想法,因为推理是从前提推得结论的过程,前提和结论展示了推理的方向。后面各章节涉及的各类数据推理将把推理建立在数据之间,往往与方向相关,如数据 a 推出数据 b 包含了从 a 到 b 的方向,是对某种数据联系的刻画描述。

就推理而言,只要对经典或非经典数理逻辑知识有所了解,必然想到逻辑推理涉及的各类推理形式,如经典形式推理、经典语义推理、各类非经典形式推理、非经典语义推理等。虽然各类逻辑推理之间存在差异,但经典或非经典逻辑推理具有共同的特点,即以逻辑公式作为推理的对象,推理都在公式之间展开。尽管不同的公式或公式之间遵循的推理规则确定了经典或非经典逻辑推理的不同形式,但公式的定义以及推理依托公式的展开使各类逻辑推理得到了统一。现不妨对公式进行适当的解释,公式是对数据(即研究的对象)性质的形式化描述,各类逻辑推理体现了数据性质之间的因果联系。例如,考虑这样的推理:如果天下雨,则地面湿。此例展示了简单且被人们接受的推理因果关系,其中天、雨和地面是涉及的对象,可视为数据。“天下雨”和“地面湿”描述了这些数据的性质,所以“如果天下雨,则地面湿”是数据性质之间的推理。数理逻辑涉及的各类推理都依托公式而展开,描述了数据性质之间的因果联系或蕴含关系。

然而,下述各章节讨论的数据推理不是数据性质之间的相互推出,如性质“天下雨”可以推出性质“地面湿”这种刻画数据性质的推理不是本书讨论的课题。本书的讨论将把推理直接建立在数据之间,将围绕“张三推出李四”、“企业 a 推出企业 b ”等这种数据之间的推理展开讨论。数据之间的推理将用以描述数据之间明确或不明确的数据关联或数据联系,这显然与经典数理逻辑和非经典数理逻辑中依托公式、在数据性质之间进行推理的推理模式存在着根本的不同,数据推理将体现本书讨论的特点。

下面的各章节将围绕数据推理展开讨论,数据推理将包括不同的形式,相关的讨论将涉及粗糙数据推理、数据关联推理、决策推理与决策系统的化简、数据

的合并与数据的转换、数据推理与三支决策等问题。这些讨论将构成本书的内容,并通过数据推理的主题予以概括或统一。

如何贯穿一条主线,使粗糙数据推理、数据关联推理、决策推理与决策系统的化简、数据的合并与数据的转换、数据推理与三支决策等各类形式的数据联系在主线下连成整体,本书采用的数据处理方法是粒计算。本书后面各章节的讨论都以粒计算的数据处理内涵或数据处理思想为支撑,使得不同的数据推理形式在粒计算的处理方式下得到统一,形成整体。

什么是粒计算?这正是本章要回答的问题。接下来将对粒计算的概念及其包含的思想进行讨论,希望能够使读者对粒计算问题形成一定的认识。由于不同的研究者对粒计算的认识存在差异,所以我们对粒计算理论及其数据处理方法看法,或本书针对粒计算的介绍及讨论将体现作者的相关思想和研究手段。不过因为常与粒计算研究者接触和交流,下面的讨论将体现粒计算方法或粒计算研究的主流做法或思想理念。

1.2 粒和粒计算的直观解释

粒计算 (granular computing, GrC) 的提出已有二十多年的历史,是出于建立数据处理方法或提供数据处理思想的考虑,为问题的算法模拟、智能处理或程序设计提供理论、方法或理念上的支撑。

粒计算涉及简单与复杂问题之间的联系,其主要意图在于通过对问题的粒化分解,使复杂问题得以简化处理,体现了问题处理的思想和数据处理的对策。因此,粒计算一经提出,便得到了专家的认可和学者的关注。多年来,针对粒计算的学术会议、国际论坛、专题研讨、专题征文、基金项目、成果交流等学术活动推进了粒计算研究的进展,活跃了课题研究的氛围,取得了有意义的成果。一些学者之所以对粒计算产生兴趣,大概可以归结为这样几个方面:①它是需要深入和系统研究的课题,就这方面而言,粒计算仍可被视为新的课题;②它从整体到部分,再利用部分之间的关系、性质、组合、运算等讨论或操作,研究整体或完成数据处理的内涵思想,为复杂问题的解决提供了思路,体现了简单与复杂之间的辩证关系;③它面向问题的处理对策或应对策略易于被从事信息处理、计算智能、逻辑推理、算法化简等方向的研究者所接受,而一些学者更将其思想方法称为艺术。正是这些因素引起了许多从事计算机科学研究者的兴趣,促进了粒计算研究热情的高涨。同时针对粒计算的研究成果也不断出现,形成了粒计算研究的初步基础,支撑着粒计算研究的进一步进展。

尽管取得了一定成绩,但如果从理论体系方面考虑,粒计算的理论研究仍有待研究者的努力,对该课题感兴趣的学者都期望系统性的粒计算理论体系的出现。

不过已有的成果具有一定的学术意义,为今后的探究提供了借鉴。本章涉及的粒计算讨论主要针对粒计算的基本概念,包括什么是粒、什么是粒之间的计算、粒度的变化、数据集的粒化处理等。本章将对粒计算中这些基本问题进行讨论,希望能帮助读者对粒和粒之间的计算等概念进行理解,形成认识,从而促进粒计算研究的进展,并期待包容、系统、深入和完整的粒计算理论体系的诞生。

后面各章节的讨论将围绕数据推理展开,将涉及数据推理不同的形式。不过这些不同的数据推理均与粒计算的数据处理内涵相一致,所以后面各章节的讨论将涉及粒计算研究的具体方法,将展示粒计算研究的具体内容。本章将针对粒计算的基本概念,可看作后面工作的准备。后面的讨论将针对具体问题,是基本概念或粒计算数据处理方法的具体体现。本章的讨论具有概括性或介绍性的特点,后面章节的方法具有针对性,可视为粒计算课题的具体内容。

如果把本章和后面各章作为整体考虑,本章是后面各章的基础,后面各章建立的方法是本章的深入。本书的各个章节的整体内容将形成具有自身特点、基于数据推理的粒计算理论体系。

以下对粒和粒之间的运算进行直观的解释。实际上,很多理论体系的产生都源于最初的直观或朴素的认识,我们首先对什么是粒进行解释性的讨论。

1.2.1 直观意义下的粒

我们认为,粒计算的核心是粒的问题。什么是粒,粒的确定产生是粒之间计算或运算(即粒计算)的基础,是粒用于数据处理的支撑。因此,对粒的认识是理解粒计算的前提,是针对粒计算展开讨论的关键环节。

要认识粒计算,需要从对粒的认识开始。从字面可以看出,粒就是颗粒,是从整体中分离出的部分,这就是对粒的直观解释。

所以,粒是相对整体而言的概念。如果稍加注意,则可以看到,粒在实际当中无处不在,也是数据处理或计算机程序化过程中时常应对的问题。现不妨考虑实际中的一些例子:

(1)如果把一个大学的全体学生看作整体,则该整体中今年考取研究生的学生是整体的部分,该部分可以看作整体的粒,该粒是统计处理、程序化管理时常涉及的对象,该对象也可以看作数据,所以粒也是数据,是由考取研究生的学生(即数据)构成的另一形式的数据。

(2)如果把一幅照片上的全部像素看作整体,则像素的二进制数值等于某一值的所有像素构成整体的部分,该部分像素的集合可看作全部像素的粒。这是图像处理或图像识别时需要应对的问题。

(3)如果把某产品(如汽车、高铁或火箭等)制造产业链上的所有企业看作整体,则该产业链中从某一企业到另一企业供货渠道上的企业是整体的部分,并可

看作粒。产业链以及产业链上的供货渠道是企业管理程序化需要面对的问题。

(4)如果把一列高铁中所有的乘客看作整体,则从某一站点上车的乘客或者下车的乘客都是整体的部分,所以上车的乘客以及下车的乘客都可以看作粒,它们都可以通过自动化的购票处理系统进行统计,使人口流动得到智能化的处理。

(5)如果把国家的所有的城市看作整体,则某些城市构成的城市群(如京津冀城市群、中部地区城市群、粤港澳大湾区城市群等)是整体的部分,一个城市群可以看作粒,且在国家发展战略中具有举足轻重的地位。

(6)如果把某国家运营航行的所有民航客机看作整体,则在某一机场降落或起飞的飞机构成整体的部分,所以与某一机场相关联的飞机可以看作整体的粒,这样的粒是机场飞行管理的重要组成部分。

(7)如果把草原上所有的动物看作整体,则位于食物链上某一层的动物是所有动物的部分,该部分是整体的粒。观察和研究该粒包含的动物,对于生态变化、环境保护、生物多样、物种分类等方面的数据处理具有重要的意义。

(8)如果把某地区的地质岩层结构看作整体,则该整体地质岩层结构中的某一层是整体的部分,可看作整体的粒。该粒包含的信息对于地质岩层的结构演化、环境变迁、地壳隆起、板块碰撞等具有重要的参考价值。

上述列举了整体与部分之间联系的相关例子,利用部分构成的粒探究整体的性质是各类研究常常涉及的问题,由此产生了利用部分探究整体的研究思想,从而提出了粒计算的研究课题。

利用信息技术的手段和方法对上述问题进行程序化的处理是信息社会的必然要求,是智能信息化的发展趋势,这将涉及描述的方法或模型的建立。粒计算就是针对方法的构建产生提出的研究课题,其目的在于通过粒的产生和粒的各类操作,以及对粒的性质进行分析,建立数据处理的数学方法,支撑算法的设计,使计算机程序化处理得以实现。

近年来,国家把人工智能提到了发展战略的层面,希望通过计算机控制或智能处理完成人类承担的工作,快速高效地处置海量的数据。这必然涉及处理方法的问题,如果不能建立计算机程序化的数学方法或应对模型,由此设定算法和编制程序,计算机是不可能自身建立方法,完成相关任务的。因此,人工智能领域的发展,依然离不开处理方法的研究、建立、扩展和完善。没有理论方法和理论体系的支撑,不可能通过计算机技术完成相关的工作。所以人工智能的发展不仅仅是技术上的处理,理论方法和理论体系的支撑才是重要的核心。

粒计算就是关于建立数据处理方法的课题,既与学术研究方面的问题相关,也源于实际中整体与部分之间的联系,其内涵可概括如下。

通过整体中某些部分包含的数据、含有的信息,以及部分与部分、部分与整体之间的联系去刻画描述问题,从而达到数据处理、解决问题的目的。