

中国海洋大学信息工程中心 研究项目
青岛海大新星计算机工程中心

COMPUTATIONAL INTELLIGENCE AND SCIENTIFIC BLENDING

计算智能 与科学配方

主 编 冯天瑾 丁香乾
副主编 杨 宁 马琳涛

揭示从经验配方到科学配方之途径。
深入研究计算智能应用，极大扩展科学技术领域。



科学出版社
www.sciencep.com

TP183/63

2008

中国海洋大学信息工程中心
青岛海大新星计算机工程中心

研究项目

Computational Intelligence and Scientific Blending 计算智能与科学配方

主 编 冯天瑾 丁香乾
副主编 杨 宁 马琳涛

科 学 出 版 社

北 京

内 容 简 介

智能科学的意义在于它大大扩展了科学技术的作用范围。本书以全新视角分析如何用计算智能方法去解决原来主要凭专业人员经验解决的高度复杂的生产实践问题,审视计算智能的基本原理、概念和应用途径;重点研究模糊集表达人对配方产品的感觉品质评估,神经网络模拟感觉评估,模糊 c -均值方法和 Kohonen 网络聚类,支持向量机对配方原料的分类和相关性分析,进化计算配方寻优,神经网络、模糊信息优化处理和模型树等知识发现方法等。以卷烟和啤酒等产品配方及药物配方为例,分析智能配方系统的设计思想和开发方略。

本书为关注计算智能理论与实践的人们而写,也可用作信息科学与技术学科和食品、饮料、医药等专业教学参考书,为相关企业技术人员与管理者提供新的思路和方法。

图书在版编目 (CIP) 数据

计算智能与科学配方/冯天瑾,丁香乾主编. 杨宁,马琳涛副主编. —北京:科学出版社,2008

ISBN 978-7-03-020603-9

I. 计… II. ①冯…②丁…③杨…④马… III. 人工智能—神经网络—计算—研究 IV. TP183

中国版本图书馆 CIP 数据核字 (2007) 第 182326 号

责任编辑:侯俊琳 程欣 王新玉/责任校对:陈玉凤

责任印制:钱玉芬/封面设计:无极书装

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>

陈海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2008 年 1 月第 一 版 开本: B5 (720×1000)

2008 年 1 月第一次印刷 印张: 17 3/4

印数: 1—4 000 字数: 343 000

定价: 35.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

本书编委会

主 编	冯天瑾	丁香乾
副主编	杨 宁	马琳涛
编 委	贺 英	石 硕
	宫会丽	刘 挺
	刘丽霞	

前 言

从人类实践意义上说,智能科学的伟大意义在于,它大大地扩展了科学技术的作用范围,使越来越多的、过去主要依靠人的经验和试探方式解决的问题,得到了明显“科学化”的解决方案;算法化和信息化系统已经深入到社会经济、生产、教育、文化和生活的各个方面。

然而,智能科学应用在我国相对薄弱,企业中的工程应用尤其值得提倡。常常有人仅仅用一些实际数据验证某个算法的应用可行性,以达到发表文章的目的。然而,这种科研模式,无论对于工程应用,还是对于学术研究的深入,都不值得提倡。工程问题通常是复杂的,涉及许多学科,而且还会有许多尚未解决的问题有待深入。

本书以解决实际问题为目标,围绕着如何用计算智能方法解决原来主要凭经验解决的复杂问题,介绍和探讨计算智能基本原理、概念和应用途径。例如,研究模糊集方法表达人对配方产品的感觉品质评估,神经网络模拟感觉评估,支持向量机对配方原料的分类,模糊 c -均值方法和 Kohonen 网络聚类,进化计算配方寻优,神经网络、模糊信息优化处理和模型树方法用于相关分析和知识发现等;以卷烟和啤酒等基于人的主观感觉质量评估的产品配方设计以及药物配方为例,介绍和分析智能配方系统的设计思想和开发方略。

围绕着产品配方问题,我们的经济实体“青岛海大新星计算机工程中心”(简称“海大新星公司”)与原颐中烟草(集团)有限公司(现“山东中烟工业公司青岛卷烟厂”)合作,开展了长达8年的应用基础研究、系统开发与应用,其成果得到相关专家和国家烟草局的高度评价与重视。国家烟草局已着手推动在全国烟草行业推广应用。2006年,我们又与青岛啤酒厂启动了智能配方技术研究;同时还着手研究在药物配方领域中的应用。

以实践为依托,我们进一步认识到,没有绝对好用或处处好用的算法,也没有最佳和最完善的特征集;即便有了正确的算法选择,也未必能解决复杂的实际问题。你必须深入实际问题本身到达如此深度,能在众多的原理和方法中优先采用其中简单有效者,反复探讨实际的模式分类和知识发现的奥秘。

所以,本书深入探讨了如下问题:复杂的配方问题往日是如何处理的?“经验配方”到底是怎么回事?在深入理解计算智能原理的基础上,分析哪些

算法可用于解决配方问题？在哪个具体问题上采用哪个算法比较好？采用哪些特征量来描述配方和配方原料？算法在解决实际问题中能起多大的作用？人的经验和智能如何进一步发挥作用，从而建立实用的人机结合智能配方系统？例如，如何结合人类专家智能与机器智能，克服配方过程中未知因素和数据不确定性带来的困难？在哪个关键点上必须与特定的领域专家（如配方师、感觉评估师、工艺师等）经验相结合？如何运用学习算法进一步发掘数据中蕴藏的知识，并有效应用于配方设计？最后，由于配方是相关产品的灵魂，配方设计牵动企业整体运行，所以，还必须将智能配方系统与企业整体信息化融为一体。

有些算法的基础数学分析已有很好的教材或专著讲述，本书不作详述，仅提出相关的参考书目。由于有些产品数据分析结果涉及企业机密，书中个别地方只好采用符号来代表某些产品属性和成分，请读者谅解。

本书为关心计算智能及其应用的人而写，适合于对计算智能、配方、感觉品质评估与企业现代集成制造系统开发有兴趣的大学生、研究生、教师、科技工作者和企业技术人员与管理者。

中国海洋大学信息工程中心
青岛海大新星计算机工程中心

目 录

前言

第一章 绪论	1
第一节 什么是计算智能	1
一、什么是智能	1
二、智能科学	2
三、计算智能与人工智能	5
四、计算智能与西药快速设计	9
五、经验科学与经验的科学	11
第二节 什么是机器学习	12
一、学习：智能之核心	13
二、什么是机器学习	15
三、机器如何学习	17
第三节 如何开发智能配方系统	19
一、智能配方系统的目标	19
二、系统开发的途径	20
三、系统开发对企业的要求	21
第二章 产品配方与感觉品质评估	23
第一节 产品质量的感觉评估	23
一、品酒师	23
二、感觉系统与感觉评估	25
三、文化背景的影响	28
四、物性感知科学	29
第二节 物性的感觉测试技术	32
一、心理物理学研究	32
二、评估小组与实验室	33
三、评估方式	34
四、感觉测试面临的问题	35
第三节 配方问题的复杂性	36

一、原料成分的复杂性	37
二、生态环境影响的复杂性	40
三、中西药配方的复杂性	42
四、什么是复杂性?	44
第四节 传统计算方法面临的问题	45
一、传统数学方法的不足	45
二、计算智能技术应用动态	48
第三章 神经网络与感觉评估	50
第一节 人工感觉系统	50
一、传感器:模拟生物感觉器官	50
二、模拟感觉系统	52
第二节 人工神经网络	53
一、什么是神经网络	53
二、与其他方法的比较	56
第三节 多层感知器:通用函数逼近器	58
一、多层感知器	58
二、反向传播(BP)学习算法	60
第四节 BP算法的改进	63
一、学习参数自适应调节	63
二、激励函数参数自适应调节	65
三、学习曲线与交叉验证	66
四、正则化与惩罚项	67
五、LMBP算法	69
第五节 网络结构优化	71
一、网络剪裁	72
二、神经元模型的推广	74
第六节 模糊集与模拟感觉评估	75
一、模糊集	75
二、模糊数与神经网络	77
三、模拟评估师·图灵测试	78
第七节 径向基函数网络	80
一、径向基函数方法	80
二、RBF网络	81
第四章 知识发现与复杂相关性分析	84
第一节 统计学相关分析能做什么	85

	一、散点图	85
	二、回归与相关分析	87
	三、线性回归分析	89
	四、SPSS 统计分析软件	90
第二节	神经网络知识发现概论	92
	一、神经网络结构分析	93
	二、神经网络功能模拟	94
	三、利用阶梯样本提取知识	97
第三节	神经网络暗箱打开了吗?	100
	一、ANN 与模糊规则基的等价性	101
	二、IRIS 分类问题	103
	三、打开 ANN 暗箱	105
第四节	网络剪裁与激励函数片线性分析	106
	一、分段线性逼近	106
	二、规则生成算法	107
	三、多种方法难分高下	109
第五节	网络权值分析与内部知识提取	110
	一、MLP 分类器	110
	二、MLP 函数逼近器的内部行为	111
	三、S 型函数的半线性分析	113
第六节	信息扩散与模糊信息优化	117
	一、信息扩散原理	118
	二、模糊变换散点图	120
	三、产品属性与原料成分关系图	122
第七节	决策树归纳学习	125
	一、用实例归纳决策树	126
	二、模型树	128
	三、模型树 M5' 算法	131
	四、啤酒属性 Y 与成分的相关分析	133
	五、烟叶成分与感觉评估和烟气指标相关分析	134
	六、烟叶感觉评估指标与烟气指标预测	137
第五章	模式识别与原料分类	139
第一节	实例启动的学习	140
	一、k-近邻分类法	140
	二、距离加权最近邻算法	142

三、学习方式：被动与主动、局部与全体·····	143
第二节 模糊 c -均值聚类·····	145
一、 c -均值聚类·····	146
二、模糊 c -均值聚类·····	147
第三节 自组织映射何以能拓扑保序·····	148
一、仿生模式识别与 SOM·····	148
二、SOM 算法要点·····	151
三、两类模拟弹性力·····	152
四、弹性网的变形与保序·····	154
五、改进 SOM 用于原配料分类·····	158
第四节 MLP：模糊线性分类器·····	161
一、模糊线性判别函数 (F-LDF)·····	161
二、MLP：模糊线性分类器·····	165
三、隐层权向量的初始化·····	166
四、最优分类超平面·····	169
第六章 支持向量机方法·····	172
第一节 统计学习与支持向量机·····	172
一、如何提高学习机的推广能力·····	172
二、VC 维与结构风险最小化·····	175
三、SVM 的基本思想·····	176
第二节 线性 SVM 模式分类器·····	180
一、最优超平面与支持向量·····	180
二、线性不可分问题求解·····	183
三、SVM 模式分类器·····	184
第三节 SVM 回归分析·····	186
一、 ϵ -不敏感误差函数·····	186
二、SVM 非线性回归·····	187
三、SVM 回归分析的特点·····	190
第四节 几种算法性能比较·····	192
一、SVM 方法之优劣·····	192
二、最小二乘 SVM·····	194
三、SVM 模拟感觉评估·····	194
四、多种算法性能比较：手写体数字识别·····	196
第七章 进化计算配方寻优方法·····	199
第一节 什么是进化计算·····	199

一、生物进化的启示·····	199
二、进化算法概述·····	201
第二节 遗传算法的基本操作·····	206
一、编码（基因链码、染色体）·····	206
二、个体与群体·····	207
三、适应度函数·····	207
四、基本遗传算子·····	209
第三节 进化算法与配方寻优·····	213
一、新药的快速发现·····	213
二、卷烟配方问题·····	214
三、叶组配方设计·····	216
第四节 多种群遗传算法·····	218
一、多种群遗传算法要点·····	218
二、进化神经网络·····	219
三、实验验证·····	222
第五节 遗传算法普适性质疑·····	224
一、模式定理·····	224
二、基因组成长假设·····	226
第八章 计算智能的若干哲理·····	228
第一节 没有包治百病的药·····	228
一、没有免费的午餐定理·····	228
二、丑小鸭定理·····	230
三、利于治疗许多病的药·····	231
第二节 经验、直觉与算法·····	233
一、算法的局限性·····	233
二、先验知识的关键作用·····	235
第三节 有从特殊到特殊的直推吗·····	237
一、归纳 + 演绎的推理·····	237
二、SVM 是从样本到样本的直推吗？·····	238
第九章 人机交互智能配方系统·····	241
第一节 集体学习·····	241
一、什么是集体学习·····	241
二、分而治之的策略·····	242
第二节 与领域专家经验相结合·····	245
一、建立专家知识库·····	245

	二、人机交互的配方优化过程·····	246
第三节	多种算法集成·····	247
	一、什么是多种算法集成·····	248
	二、多算法集成的另一实例·····	250
第四节	配方系统与现代集成制造系统·····	251
	一、智能配方系统结构·····	252
	二、配方系统融入企业信息化大系统·····	258
	三、从经验配方到科学配方·····	260
参考文献	·····	263
致谢	·····	272

第一章

绪 论

所以知之在人者，谓之知。知有所合，谓之智。所以能之在人者，谓之能。能有所合，谓之能。

——荀况《荀子·正名》

从古至今，人类一直在观察、研究自己并试图用科学技术给自己制作模型。

——David Poole 等《Computational Intelligence》

第一节 什么是计算智能

生命，是浩瀚宇宙最奇妙之花朵。生命世界中，最复杂、最美好和最值得骄傲的是智能。智能科学与技术乃现代科技之热点，未来科技发展之关键，全世界瞩目。

一、什么是智能

通常，我们将人的能力分为体力和智力，或体能和智能。智能，或智力、智慧，对应的英文词为 intelligence（在台湾译为“智慧”）。像给许多的概念下定义那样，给智能下一个简明定义也很难。

《辞海》（1979 年版）给出了一个具有代表性的定义：

智力，通常叫“智慧”。指人认识客观事物并运用知识解决实际问题的能力。集中表现在反映客观事物深刻、正确、完全程度上和应用知识解决实际问题的速度和质量上，往往通过观察、记忆、想像、思考、判断等表现出来。它在掌握人类知识经验和从实践活动中发展，但又不等同于知识和实践。它是先天素质、社会历史遗产与教育的影响以及个人努力与实践三方面因素相互作用的产物。

问题出现了，这类定义谈的只是“人类智力”。我们没有发现古代学者论述过人以外的生物的智能现象，他们理所当然地认为人是“万物之灵”。但是，只有人类有智力吗？猿人、猴子有没有智能？其他生物呢？当代，人们已经承认有脑的动物有智能，并对有些动物的智力之高超惊讶不已。还有植物学家介绍，有一种“金合欢树”在遇到动物攻击的时候，会产生一种具有难闻味道的丹宁酸，以阻止动物啃食自己。同时，被啃过树的树叶会释放一种气体，周围其他的金合欢树能识别这种气体，从而也释放出丹宁酸，使得动物不敢来犯。类似的例子还很多。所以，一些植物学家说，尽管植物行动迟缓，它们实际上比我们通常认为的能干得多。它们具有一定的识别事物和解决问题的能力，即具有一定的智力。

从微观来看，20 世纪中期有科学探讨：脑内部的神经元和它们组成的网络是怎么工作的？是否具有智能？如果它们没有智能，脑的整体智能从何而来？还有大宏观的生命进化，生物遗传、变易与进化机理是否意味着一种智力呢？显然，我们很难用前面《辞海》给出的智力定义中所言，说这些都是“通过观察、记忆、想像、思考、判断等表现出来”的能力，不能这样描述和理解神经元和神经网络学习能力。我们后面还要详细讨论这些问题。

面对越来越丰富的内涵，目前学术界对智能的定义无法统一。我们觉得比较好的定义可表述为：

智能是一种成功地适应环境、有效地找到问题的满意解的能力。

这个定义既适合于定义脑的宏观智能，甚至植物的某种智能，也可包容脑微观组织和神经元的智能，还可以描述生命进化这样的大系统以及免疫系统的功能，直到机器的智能。这个定义里用的词“问题”，既可以是环境给生物带来的问题，也可以是神经网络面对输入输出（环境）信息的约束。当然，它更可以是人类或者机器面临的问题。总之，是“智能体”面临的问题。

二、智能科学

现代科学承认，尽管我们制造了很高级的机器，计算机每秒能完成数万亿次运算，速度是人脑的 10^{12} 倍，全世界的人一起计算速度总和的 100 倍；100G 的计算机硬盘可以存放 10 万本书的信息。但是，就感觉、运动协调和解决问题的综合能力而言，传统计算机的智力不及一只蚊子！生物学家和仿生学者早就发现了大量事实，足以证明生物神经计算（感知、识别、决策和自身自适应控制）的高超能力（图 1.1）。



图 1.1 蚊子：能轻巧地落在水面上；能从水面上垂直起飞（右）；能避开雨点飞行

智能科学在一定程度上是人类不断研究自己的成果，是在不断探索“生物和人的智慧是怎么一回事”和“如何建造智能实体”这两个问题之中成长的。

首先得到较深入研究的是思维逻辑。我国春秋（公元前 770 ~ 前 476）和战国（公元前 403 ~ 前 221）时期，各诸侯国争霸，也是多种思想、学派自由发展与竞争的时代。由于政治斗争、学术辩论的需要，首先要把当时较为混乱的关于名词（名）与其所代表的事物（实）的关系（名实关系）搞清楚。继而是辩术（思辨规律）研究，促进了所谓名辩思想（逻辑学思想）大发展。古希腊研究逻辑学的突出代表人物是亚里士多德（Aristotle，公元前 384 ~ 前 322）。他试图“编纂理性思维”，将其规则形式化为逻辑学（形式逻辑），建立了一种语义结构理论和一个复杂的逻辑规则系统。他认为，这些思维规则可以用来管理大脑思维运作，人们似乎能用这个系统严格地推出所有的结论。

遵循这条道路，不断有更深入的哲学和科学成果出现。英国著名唯物论哲学家霍布斯（T. Hobbes，1588 ~ 1679）说：“一切思维不过是计算（即加与减）。”最先将思维与计算相联系，落实并尝试建立数学逻辑的是德国哲学家、数学家和物理学家莱布尼茨（G. W. Leibniz，1646 ~ 1716）。莱布尼茨坚信，基于一种统一的科学语言——符号化方法，可建立“普遍逻辑”和“逻辑演算”，世界上一切都可以这样解释清楚。

法国唯物主义哲学家、医生拉·美特利（J. O. de La Mettrie，1709 ~ 1751）以他的名著喊出了响亮的口号《人是机器》（1747）。他说：

宇宙中存在的只有物质组织，而人则是其中最完善的一种。

人的身体是一台钟表，不过，这是一台巨大的、极其精细、极其巧妙的钟表。……是黑胆、苦胆、痰汁和血液等体液按其性质、多寡与配合方式的不同，使得人与别人不同。

现代科学用神经介质多巴胺、内啡肽和基因发现，极大地丰富与发展了

拉·美特利的这一思想，带来了极为辉煌的成果。自“人是机器”的唯物主义口号提出以来，一直是智能模拟、机器思维和仿生学的一面旗帜。当今科学界，“人、动物和计算机都是信息处理系统”、“人是自动机”、“意识是脑神经的活动”等，都是拉·美特利口号的变体。

1847年，爱尔兰数学家乔治·布尔（G. Boole, 1815~1864）出版了名著《逻辑的数学分析》，提出了逻辑代数，为逻辑推理建立了形式语言。他的《思想规律的研究》（1854）发挥了一个重要思想：“符号语言与运算可用来表示任何事物。”作为数理逻辑第一人，他使逻辑学由哲学变成了数学。在逻辑与符号体系方面，德国数学家、哲学家和谓词逻辑创始人弗雷格（G. Frege, 1848~1925），英国大数学家罗素（B. Russell, 1872~1970）与他的老师怀特海（A. N. Whitehead, 1861~1947）等，做出了进一步理论研究。

现代人工智能符号论学派的代表人物纽厄尔（A. Newell, 1927~1992）和西蒙（H. A. Simon, 1916~2001，有人译为司马贺）坚定地沿着“推理即符号处理”之道前进。他们坚信：

人是用符号处理来解决问题的，无论人们是在论述、说话、书写、学习，还是发明，都是如此。

符号系统，是模拟人通过逻辑推理解决问题的思路最成功的方法。

语言是通向人类心灵的一扇窗户，而语言就是符号的传递。人类群体或社会的智能更是依靠语言和其他信号传递。符号方法可以描述事物的目标状态和当前状态，而从当前状态演变到目标状态的过程序列就是程序。反过来，这些程序有可能用于对人做心理实验，进而创建可测试的关于人脑工作方式的理论。

其实，这只是智能学发展中具有悠久历史的一条道路：研究辩论术→语言规则→逻辑→符号逻辑·数理逻辑→问题求解→物理符号系统（人脑和电脑）。它就是基于人类理性和逻辑学、被称为“符号主义”（符号论）的路线。

另外一条研究道路起于脑内部结构研究。在那样远古年代，被后人誉为医学之父的希腊医生、教师希波克拉底（Hippocrates, 公元前460~前370）就开始了探究脑是干什么的伟大研究，得出了大脑是人类意识与感情源地的结论。意大利解剖学家、病理学家高尔基（C. Golgi, 1844~1926）和西班牙解剖学家卡加尔（R. Cajal, 1852~1934）于1872年创立的神经元学说（Neuron Doctrine），带领人们进入了理解脑工作机理的大门，奠定了现代脑科学——神经科学发展的基础。

美国生理学家麦克卡洛（W. S. McCulloch）和数学家匹茨（W. Pitts）的

长期合作研究,把脑神经系统研究、数理逻辑思想以及图灵的计算理论相结合,提出了以后被称为 MP 模型的脑模型(1943),这一次科学飞跃,可视为计算智能的起步。美国心理学家赫伯(D. O. Hebb, 1904 ~ 1985)发现了在环境刺激下神经元之间连接的演进方式,即所谓神经元的赫伯学习规则。从而展开了研究脑的结构→脑皮层分区→神经元·神经学习→神经网络重组,从“微观角度”研究生物神经系统到人工神经网络的“连接主义”(连接论)路线。

1956 年夏,在美国 Dartmouth 大学举办了一个异乎寻常、长达 2 个月的研讨会。会议围绕着“机器如何模拟人类智能”这个中心问题,认真、热烈、深入地讨论了符号系统、神经网络、机器学习与自动化等领域的基本思路。与会者接受了麦卡锡设计的名字:人工智能(artificial intelligence, AI)。因此,这个研讨会被公认为是人工智能学科诞生的标志。此后仅仅两个月,即 1956 年 9 月,在麻省理工学院召开了另一个研讨会,人们把它视为认知科学的起点(Russell and Norvig, 2004)。心理学家还认为,它是认知心理学诞生之日,认知科学是认知心理学的研究方法之一(艾森克和基恩, 2000)。

智能科学还远不只上述内容。由于生物智能的广博,智能科学也自然丰富多彩。

人们还进一步从大自然获得了灵感。20 世纪 60 年代,美国加利福尼亚大学工程科学博士弗格(D. B. Fogel)在研究自动机时发展了“进化规划”方法。密歇根(Michigan)大学心理学、计算机科学教授霍兰(J. H. Holland)和他的合作者们,从解释生物的复杂适应过程,到模拟生物进化机制,自然计算(进化计算、遗传算法)跨出了关键的一步。1986 年,物理学家、复杂系统与混沌理论家法默(J. D. Farmer)等人发明了人工免疫算法。1987 年 9 月的第一次国际人工生命研讨会,确立了以数学家兰顿(C. G. Langton)为代表的“人工生命”理论。这种“算法+自然机制”的模式已得到了丰富的展开。

三、计算智能与人工智能

由于智能学的内容极为丰富,发展极为迅速,学术名词繁多,许多问题尚无定论。自从 17 世纪霍布斯明确地将思维与计算相联系以后约 300 年,图灵开创了计算理论与人工智能一体性研究,他那有划时代意义的论文“计算机器与智能”(1950),可以说是“计算智能”(computational intelligence, CI)术语的来由。虽然术语“人工智能”与“计算智能”的含义相近,仍然有两大原因使得这两个术语分野。