

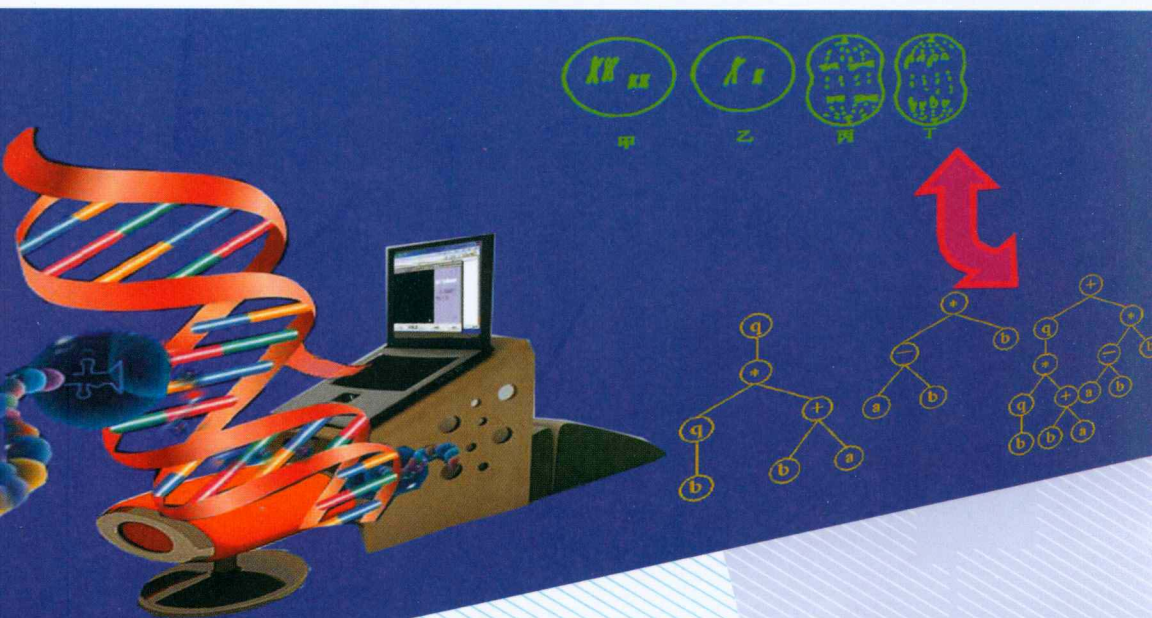


智能

科/学/技/术/著/作/丛/书

# 基因表达式编程算法原理与应用

元昌安 彭昱忠 覃晓 石亚冰 蔡宏果 编著



科学出版社

[www.sciencep.com](http://www.sciencep.com)

智能科学技术著作丛书

# 基因表达式编程算法 原理与应用

元昌安 彭昱忠 覃 晓 石亚冰 蔡宏果 编著

科 学 出 版 社

北 京

## 内 容 简 介

基因表达式编程是在继承和发展遗传算法与遗传编程优点的基础上发展起来的遗传计算家族中的新成员,是借鉴生物遗传的基因表达规律提出的搜索和知识发现新技术。研究已表明,它在很多领域具有比遗传算法和遗传编程更强的解决问题的能力。全书共分四部分(14章)。第一部分介绍了基因表达式编程产生的相关背景;第二部分系统地阐述了基因表达式编程的基本概念、基本原理、基本算法和理论分析;第三部分着重讨论了基因表达式编程算法的拓展和改进;第四部分重点介绍了基因表达式编程在知识发现等多个领域中的典型应用。

本书可作为高等院校计算机科学与技术、软件工程等相关专业高年级本科生的选修教材和数据挖掘、智能计算等方向研究生参考教材,特别是可作为所有拟对基因表达式编程技术进行深入研究或借助基因表达式编程技术进行应用研究的学者、工程师们的参考用书。

### 图书在版编目(CIP)数据

基因表达式编程算法原理与应用/元昌安等编著. —北京:科学出版社,2010

(智能科学技术著作丛书)

ISBN 978-7-03-028698-7

I. 基… II. 元… III. 基因表达-程序设计-算法-理论 IV. ①TP311  
②TP18

中国版本图书馆 CIP 数据核字 (2010) 第 161787 号

责任编辑:王志欣 张艳芬 / 责任校对:刘亚琦

责任印制:赵 博 / 封面设计:耕者

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>

源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

\*

2010 年 8 月第 一 版 开本: B5 (720×1000)

2010 年 8 月第一次印刷 印张: 23 1/2

印数: 1—3 000 字数: 449 000

定价: 68.00 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 蔡自兴 孙增圻 谭 民

秘 书 长：韩力群

编 委：（按姓氏汉语拼音排序）

蔡庆生（中国科学技术大学）

杜军平（北京邮电大学）

何华灿（西北工业大学）

黄河燕（中国科学院计算语言研究所）

焦李成（西安电子科技大学）

刘 宏（北京大学）

秦世引（北京航空航天大学）

阮秋琦（北京交通大学）

孙增圻（清华大学）

涂序彦（北京科技大学）

王家钦（清华大学）

吴文俊（中国科学院系统科学研究所）

尹怡欣（北京科技大学）

张琴珠（华东师范大学）

庄越挺（浙江大学）

蔡自兴（中南大学）

韩力群（北京工商大学）

何 清（中国科学院计算技术研究所）

黄心汉（华中科技大学）

李祖枢（重庆大学）

刘 清（南昌大学）

邱玉辉（西南师范大学）

史忠植（中国科学院计算技术研究所）

谭 民（中国科学院自动化研究所）

王国胤（重庆邮电学院）

王万森（首都师范大学）

杨义先（北京邮电大学）

于洪珍（中国矿业大学）

钟义信（北京邮电大学）

## 《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。

1956年,“人工智能”学科诞生,50年来,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

## 序

十年前,研究进化生物学的 Ferreira 博士融合了遗传算法的两个传统领域——遗传算法和遗传编程,提出了遗传计算家族中的新算法——基因表达式编程。

遗传算法的染色体是线性定长串,旨在通过简单编码解决简单问题;遗传编程的染色体是不定长非线性树结构,旨在利用复杂编码解决复杂问题;基因表达式编程用定长线性串表达非线性树结构,是利用简单编码解决复杂问题。基因表达式编程一方面继承了遗传算法的刚性、规矩和快速,另一方面继承了遗传编程的柔性易变和多能。

2000 年,当 Ferreira 报告基因表达式编程的效率比老的遗传编程系统高出  $10^2 \sim 6 \times 10^6$  倍时,学术界还将信将疑。在一片怀疑声中,网页设计和图形软件开发者 José Simas 对基因表达式编程给予了高度评价,并同 Ferreira 合作创办了基因表达式编程公司,开发了基于基因表达式编程的软件,同时在理论、工程和系统三方面对传统思维的发起了冲击。2001 年, Ferreira 在 *Complex Systems* 杂志第 2 期发表了其原创性论文 *Gene Expression Programming: A New Adaptive Algorithm for Solving Problems*; 2002 年, Ferreira 出版了有关基因表达式编程的第一本专著 *Gene Expression Programming: Mathematical Modeling by An Artificial Intelligence*, 该公司还建立了关于基因表达式编程的最新成果的网站(<http://www.gene-expression-programming.com/>基因表达式编程 Biblio. asp)。

当基因表达式编程在地球村那一半刚一出现,就在中国找到了知音。四川大学数据库与知识工程研究所的研究团队,特别是左勃博士,当 Ferreira 原创性论文在网上出现 10 多天,尚未正式发表时,以特有的兴趣和学术敏感,捕捉了机会,将基因表达式编程第一时间引入国内,该团队连续得到两个国家自然科学基金和一个博士点基金,开展了深入的研究。在国际会议 WAIM 2002 上,该团队发表了国内第一篇关于基因表达式编程的研究论文。在文中用数学归纳法严格地证明了被 Ferreira 直接采用而未加以证明的收敛定理:“对任意良好定义的基因, Ferreira 的解码过程(算法)总能成功返回到对应的表达式树的根节点”,这一定理在理论上证明了 Ferreira 的解码算法总会成功收敛,扫除了理论上的障碍,拉开了国内关于基因表达式编程研究与应用的序幕。

正是上述这个收敛定理,保证了基因表达式编程中,在残酷的遗传操作(如切头、去尾、交叉)之下,基因仍然是“活的”。而在传统的遗传编程和遗传算法中,

在如此残酷的遗传操作下,绝大多数基因都死去了,如编程不可解析或无意义表达式,或者在解析表达式时引用空指针的被指物(可能导致死机),或者出现分母为 0 以及在实数领域中负数开方等;大量的死基因浪费了计算资源。这一定理也解释了为什么基因表达式编程的效率比原始的遗传编程系统高出几个数量级。在今天的理论和实践上的成功面前,已经不再有人怀疑基因表达式编程的高效率。

资料表明,目前国内从事基因表达式编程研究的学术队伍不断壮大,学术论文迅速增多。至今,关于基因表达式编程的硕士、博士学位论文已经分布在包括四川大学、武汉大学、华中科技大学、兰州大学、华南理工大学、中国地质大学、华北电力大学、江西理工大学、广西师范学院在内的数十所高校,专业涉及十余个,这充分反映了国内关于基因表达式编程的研究在深入进行,基因表达式编程在其他学科的应用也在飞速扩展。

基因表达式编程是遗传算法家族中的新成员,其总体思路 and 传统遗传算法相同。设要在一批数据中找出描述潜在规律的公式(即优化目标),则从随机初始化的公式群出发,从中选出较好的公式作染色体,通过杂交或变异,对子女进行适应度评估,选出较好的,再进行杂交或变异,再评估;如此迭代,子子孙孙没有穷尽,在代代传承中,继承和遗传是主流;但有两种契机可加快进化进程,或者是种群内有优秀个体横空出世(变异);或者是引进了外部优秀基因,使种群多样化。在基因表达式编程中,遗传和变异都在基因上进行,收敛定理保证了基因在遗操作下不会死去,保证了一定收敛,且比遗传算法和遗传编程快几个数量级。

随着基因表达式编程向其他学科的渗透,目前国内还没有一本系统介绍基因表达式编程的中文专著。借此契机,元昌安和他的团队系统整理了基因表达式编程的研究成果。全书有浅、实、新、兼的特色。浅——入门知识浅、语言浅显易懂,循序渐进地表达知识内容;实——理论和实际相结合,算法原理与实践相结合;新——整个基因表达式编程的内容新颖,有许多内容是四川大学“数据库与知识工程”研究团队和广西师范学院“科学计算与智能信息处理”研究团队的研究成果;兼——点面兼顾,目录分类细致,方便不同类型读者的快速查阅,附有标准基因表达式编程的代码,有利于初学者提高学习效率,同时能很快过渡到研究阶段。

基因表达式编程是遗传计算家族在新世纪之初出现的新成员,是数据挖掘特别是公式发现的利器。但愿这本著作的出版能对从事该领域的科研人员起到抛砖引玉的效果,能对刚刚涉入该领域的初学者起到指点迷津的作用。

唐常杰

2010 年 7 月

# 前 言

基因表达式编程是借鉴生物遗传的基因表达规律提出的知识发现新技术, Ferreira 的原创思想萌芽于 1999 年,2000 年底在网上开始有报道,2001 年底正式发表,并冠以副标题 *Complete Reference for the First GEP Paper* 以强调其原创性。

实验表明,基因表达式编程的效率比原始的遗传编程系统竟然高出 100~60000 倍。作为进化生物学家兼计算机科学家,Ferreira 用生物学逻辑来解释:“基因表达式编程跨越了表现型的极限。这意味着由 GEP 进化的复杂的计算机程序(表现型)被完全编码在一个固定长度的简单字符串里面(染色体或者基因型)”。

Ferreira 于 2001 年发表了关于 GEP 的第一篇论文,于 2002 年出版了关于 GEP 的第一本专著,从此开创了 GEP 研究与应用的先河。随后,Ferreira 又发表了十余篇相关学术论文,并于 2006 年出版了专著的第二版。国际上关于 GEP 的研究也呈蓬勃发展之势。

国内关于 GEP 的研究几乎与国际同步,四川大学数据库与知识工程研究所唐常杰教授和左昞博士以高度的学术敏感性,于 2001 年将 GEP 的研究引入了国内。国内关于 GEP 研究的学术队伍遍布大江南北的数十所高校,并正向多个专业领域渗透,反映出 GEP 广阔的应用前景。

关于 GEP 当前研究的着眼点主要集中在以下四个方面:

(1) 用实验验证 GEP 各种遗传策略的效率,即对 GEP 进行“内部比较”;主要成果反映在 Ferreira 的学术论文和专著中。

(2) 针对解决问题的特性,提出了一系列新概念和新算法。

(3) 将 GEP 技术应用于其他学科的研究中,探究多学科交叉融合方法及其应用。

(4) 对 GEP 的形式化、收敛性等进行理论分析和研究,旨在探讨 GEP 的机理,在理论上保证 GEP 方法的可靠性和可行性。

目前,国内还没有一本系统介绍 GEP 的中文专著。系统整理 GEP 的研究成果,出版一本中文专著供国内相关研究者参考,是一件重要而有意义的工作。

本书分为四部分(共 14 章)。其中第一、二两部分是入门和基础,第三、四两部分是深入和提高。

第一部分(第 1,2 章)背景概述篇:介绍了基因表达式编程的相关背景,概述了基因表达式编程产生的相关生物学理论和背景、发展史和特点。初学者通过本部

分的学习可以了解进化算法的基本知识,了解基因表达式编程的生物学背景和算法概况。

第二部分(第3~5章)基本算法与理论篇:介绍了基因表达式编程的基本算法与理论,具体包括基因表达式编程算法的基本概念、基本操作、基本结构及实现原理、算法的设计和步骤等;分析了算法的可靠性、完备性和收敛性。读者通过学习本部分可以了解 GEP 算法及其内在机理。

第三部分(第6~9章)算法拓展与改进篇:介绍了基因表达式编程算法的拓展、改进和改进成果。具体包括个体的评价策略及其计算方法;算法中常数的创建和表示方法;优秀基因片段的保留和传承;个体的表示和组织结构;种群的生成和组织管理;遗传操作的改进和经典遗传行为的干预;基因表达式编程的并行化方法和实现;基因表达式编程与其他智能算法的融合等。

第四部分(第10~14章)应用研究篇:介绍了 GEP 在函数挖掘、预测、关联规则、分类聚类等研究领域中的典型应用。具体包括 GEP 在函数挖掘中的应用、挖掘递归函数、挖掘中医方证关系等;时间序列和数据流预测模型、灾情分析、多层关联规则挖掘、分类和聚类、特征自动抽取、文本分类;在数据压缩、演化硬件研究中的应用。

本书编写时注重如下特色:语言浅显易懂、循序渐进、结合实例、学以致用;目录细致、便于查阅;附有标准 GEP 的代码,入门容易。

本书由元昌安、彭昱忠、覃晓、石亚冰、蔡宏果编著,其中第一部分由石亚冰撰写;第二部分由元昌安、石亚冰、彭昱忠、蔡宏果撰写;第三部分由彭昱忠、蔡宏果撰写;第四部分由覃晓撰写;附录由蔡宏果完成。王艳为封面设计做出了重要贡献,并参与了校对工作。元昌安对全书进行了统稿。

本书内容主要来自四川大学数据库与知识工程和广西师范学院科学计算与智能信息处理研究团队的研究成果,在此向四川大学数据库与知识工程研究团队致以崇高的敬意和由衷地感谢!向广西师范学院科学计算与智能信息处理研究团队的同仁们,以及为本书的出版做出贡献的邓松、廖剑平、饶元、磨高尚、王艳、丁超、刘海涛、李桂来、宋叶俊、陶俊剑、张增银、王文栋等表示感谢!同时,本书中还引用了其他许多学者的研究成果,在此对他们表示衷心的感谢!

本书由国家自然科学基金(60763012)、广西高校重点实验室及人才小高地创新团队科学计算与智能信息处理项目经费、广西师范学院学术著作出版基金和硕士点建设费资助出版。

书中不足之处在所难免,恳请读者批评指正。

作 者

2010年7月

# 目 录

《智能科学技术著作丛书》序

序

前言

## 第一部分 背景概述篇

|                      |    |
|----------------------|----|
| 第 1 章 最优化问题          | 3  |
| 1.1 最优化问题简述          | 3  |
| 1.2 最优化问题的求解方法       | 3  |
| 1.2.1 求解最优化问题的数学方法   | 4  |
| 1.2.2 进化计算求解方法       | 4  |
| 1.2.3 无免费午餐定理        | 5  |
| 参考文献                 | 5  |
| 第 2 章 生物进化与进化计算      | 6  |
| 2.1 从进化论到进化计算        | 6  |
| 2.1.1 生物进化论概要        | 6  |
| 2.1.2 遗传算法           | 7  |
| 2.1.3 进化策略           | 8  |
| 2.1.4 进化规划           | 8  |
| 2.1.5 遗传编程           | 10 |
| 2.2 广义的进化计算          | 11 |
| 2.2.1 进化计算的本质        | 11 |
| 2.2.2 进化算法的一般框架      | 11 |
| 2.3 生物的基因表达          | 12 |
| 2.3.1 DNA            | 12 |
| 2.3.2 RNA            | 13 |
| 2.3.3 蛋白质            | 14 |
| 2.3.4 基因表达过程         | 14 |
| 2.4 从生物的基因表达到基因表达式编程 | 15 |
| 2.4.1 GEP 的发展历史      | 15 |

|                        |    |
|------------------------|----|
| 2.4.2 GEP 的研究和应用 ..... | 15 |
| 2.4.3 GEP 的特点 .....    | 15 |
| 参考文献 .....             | 16 |

## 第二部分 基本算法与理论篇

|                                 |    |
|---------------------------------|----|
| <b>第 3 章 GEP 算法基础</b> .....     | 21 |
| 3.1 GEP 的基本要素 .....             | 21 |
| 3.1.1 开放读码框架和基因 .....           | 22 |
| 3.1.2 GEP 中的基因 .....            | 24 |
| 3.1.3 多基因染色体 .....              | 27 |
| 3.1.4 子表达式树的相互作用 .....          | 28 |
| 3.2 基本的遗传操作 .....               | 30 |
| 3.3 GEP 基本算法 .....              | 32 |
| 3.4 适应度函数的选择和评估 .....           | 33 |
| 3.5 标准 GEP 的选择策略 .....          | 37 |
| 3.6 简单的函数发现实例 .....             | 40 |
| 参考文献 .....                      | 45 |
| <b>第 4 章 GEP 的理论分析</b> .....    | 46 |
| 4.1 形式化定义 .....                 | 46 |
| 4.1.1 终结符和函数 .....              | 46 |
| 4.1.2 概念的形式化定义 .....            | 48 |
| 4.2 基因编码表达的可靠性和完备性分析 .....      | 50 |
| 4.3 GEP 的马尔可夫链分析 .....          | 51 |
| 4.3.1 马尔可夫链相关概念 .....           | 51 |
| 4.3.2 GEP 马尔可夫收敛定理 .....        | 53 |
| 4.4 GEP 的依概率收敛分析 .....          | 54 |
| 4.4.1 问题背景及相关定义 .....           | 54 |
| 4.4.2 基于 GEP 的函数发现依概率收敛定理 ..... | 55 |
| 参考文献 .....                      | 58 |
| <b>第 5 章 GEP 的进化过程分析</b> .....  | 60 |
| 5.1 个体与群体结构分析 .....             | 60 |
| 5.1.1 个体结构分析 .....              | 60 |
| 5.1.2 群体结构分析 .....              | 62 |
| 5.2 遗传算子性能分析 .....              | 64 |

|                      |    |
|----------------------|----|
| 5.2.1 选择 .....       | 65 |
| 5.2.2 变异 .....       | 67 |
| 5.2.3 移位 .....       | 67 |
| 5.2.4 重组(交叉) .....   | 68 |
| 5.2.5 总结 .....       | 69 |
| 5.3 个体的建筑块实验分析 ..... | 69 |
| 参考文献 .....           | 72 |

### 第三部分 算法拓展与改进篇

|                              |     |
|------------------------------|-----|
| 第 6 章 GEP 个体与种群组织的拓展 .....   | 75  |
| 6.1 个体评价 .....               | 75  |
| 6.1.1 经典领域的适应度函数设计 .....     | 75  |
| 6.1.2 动态适应度 .....            | 78  |
| 6.2 基因的评估 .....              | 82  |
| 6.2.1 GRM 算法 .....           | 84  |
| 6.2.2 Kquick 算法 .....        | 85  |
| 6.2.3 GPED 算法 .....          | 86  |
| 6.2.4 SGDE-GEP 算法 .....      | 88  |
| 6.2.5 基于 Scale 的基因评估算法 ..... | 94  |
| 6.3 常数问题 .....               | 101 |
| 6.3.1 三类主要方法 .....           | 101 |
| 6.3.2 常数创建方法典型改进 .....       | 106 |
| 6.4 优秀基因片段的传承 .....          | 110 |
| 6.4.1 基于个体结构的实现方法 .....      | 110 |
| 6.4.2 基于编码映射的实现方法 .....      | 113 |
| 6.5 染色体的组织 .....             | 115 |
| 6.5.1 多层染色体 .....            | 116 |
| 6.5.2 基因重叠染色体 .....          | 123 |
| 6.5.3 带身部结构的染色体 .....        | 133 |
| 6.5.4 DAG 染色体 .....          | 134 |
| 6.6 种群的组织 .....              | 136 |
| 6.6.1 改进的初始种群生成策略 .....      | 136 |
| 6.6.2 种群多样性繁殖策略 .....        | 143 |
| 6.6.3 变种群规模策略 .....          | 147 |

|                                   |     |
|-----------------------------------|-----|
| 6.6.4 其他种群组织策略 .....              | 157 |
| 参考文献 .....                        | 158 |
| <b>第7章 GEP 遗传操作与行为干预的拓展</b> ..... | 160 |
| 7.1 遗传操作拓展 .....                  | 160 |
| 7.1.1 简单分组的遗传操作 .....             | 160 |
| 7.1.2 残差制导进化的 GEP 算法 .....        | 161 |
| 7.1.3 自适应 GEP 算法 .....            | 164 |
| 7.2 GEP 的转基因策略 .....              | 172 |
| 7.3 GEP 的回溯策略 .....               | 179 |
| 参考文献 .....                        | 183 |
| <b>第8章 并行 GEP 算法</b> .....        | 184 |
| 8.1 并行遗传算法的实现方案 .....             | 184 |
| 8.2 并行 GEP 的设计 .....              | 188 |
| 8.2.1 同步分布式并行 GEP 算法 .....        | 188 |
| 8.2.2 异步分布式并行 GEP 算法 .....        | 189 |
| 参考文献 .....                        | 190 |
| <b>第9章 GEP 与其他算法的融合</b> .....     | 191 |
| 9.1 GEP 与人工神经网络的融合 .....          | 191 |
| 9.1.1 基于 GEP 和神经网络的属性约简分类算法 ..... | 191 |
| 9.1.2 基于 GEP 优化的 RBF 神经网络算法 ..... | 199 |
| 9.2 GEP 与模拟退火的融合 .....            | 205 |
| 9.3 GEP 与禁忌搜索的融合 .....            | 211 |
| 9.4 GEP 与隐马尔可夫模型的融合 .....         | 219 |
| 参考文献 .....                        | 224 |

## 第四部分 应用研究篇

|                                      |     |
|--------------------------------------|-----|
| <b>第10章 GEP 在函数挖掘中的应用</b> .....      | 229 |
| 10.1 用 GEP 进行函数发现 .....              | 229 |
| 10.1.1 一致表达式的发现 .....                | 229 |
| 10.1.2 分域表达式挖掘 .....                 | 232 |
| 10.1.3 复杂度分析 .....                   | 235 |
| 10.1.4 实验与分析 .....                   | 236 |
| 10.2 用 GEP 挖掘递归函数 .....              | 239 |
| 10.2.1 递归函数挖掘算法 GEP-RecurMiner ..... | 239 |

|                                   |     |
|-----------------------------------|-----|
| 10.2.2 动态进化策略 .....               | 242 |
| 10.2.3 实验与分析 .....                | 243 |
| 10.3 用 GEP 挖掘中医方证关系 .....         | 245 |
| 10.3.1 中医学中的方证关系 .....            | 245 |
| 10.3.2 简单方证关系分析 .....             | 246 |
| 10.3.3 症状组主次分离指针编码 .....          | 247 |
| 10.3.4 中医学中的症状组适应度函数设计 .....      | 250 |
| 10.3.5 基于基因表达式的方证关系函数发现算法 .....   | 252 |
| 10.3.6 实验与分析 .....                | 252 |
| 参考文献 .....                        | 258 |
| <b>第 11 章 GEP 在预测中的应用</b> .....   | 260 |
| 11.1 GEP 在时间序列预测中的应用 .....        | 260 |
| 11.1.1 Fibonacci 加权滑动窗口预测法 .....  | 261 |
| 11.1.2 差分平均预测法 .....              | 263 |
| 11.1.3 实验与分析 .....                | 265 |
| 11.2 GEP 在数据流预测模型中的应用 .....       | 270 |
| 11.2.1 数据流的基本概念和定义 .....          | 271 |
| 11.2.2 基于 GEP 数据流挖掘算法 .....       | 272 |
| 11.2.3 多数据流的预测模型和算法介绍 .....       | 272 |
| 11.2.4 实验与分析 .....                | 275 |
| 11.3 GEP 在灾情分析及预测中的应用 .....       | 278 |
| 11.3.1 洪灾灾情预测 .....               | 278 |
| 11.3.2 基于折半查找的最优属性约简算法 .....      | 278 |
| 11.3.3 基于属性约简的 GEP 洪灾灾情预测算法 ..... | 279 |
| 11.3.4 实验与分析 .....                | 280 |
| 参考文献 .....                        | 281 |
| <b>第 12 章 GEP 在规则挖掘中的应用</b> ..... | 283 |
| 12.1 GEP 在关联规则挖掘中的应用 .....        | 283 |
| 12.1.1 关于关联规则挖掘 .....             | 283 |
| 12.1.2 谓词关联规则概念 .....             | 284 |
| 12.1.3 基因和适应度函数设计 .....           | 285 |
| 12.1.4 挖掘谓词关联规则 .....             | 287 |
| 12.1.5 实验与分析 .....                | 288 |
| 12.2 基于 GEP 的多层关联规则挖掘算法及其应用 ..... | 289 |
| 12.2.1 多层关联规则挖掘 .....             | 290 |

|                                    |     |
|------------------------------------|-----|
| 12.2.2 基于泛化的抽象频繁项集挖掘 .....         | 290 |
| 12.2.3 挖掘多层关联规则算法 .....            | 290 |
| 12.2.4 实验与分析 .....                 | 293 |
| 参考文献 .....                         | 295 |
| <b>第 13 章 GEP 在分类和聚类中的应用</b> ..... | 296 |
| 13.1 基于 GEP 的自动聚类方法 .....          | 296 |
| 13.1.1 染色体编码 .....                 | 296 |
| 13.1.2 聚类表达式树的编码和解码 .....          | 297 |
| 13.1.3 适应度计算 .....                 | 299 |
| 13.1.4 遗传算子 .....                  | 299 |
| 13.1.5 基于 GEP 的聚类算法 .....          | 299 |
| 13.1.6 自动合并整合簇算法 .....             | 300 |
| 13.1.7 实验与分析 .....                 | 301 |
| 13.2 代价敏感的 GEP 分类算法 .....          | 304 |
| 13.2.1 代价敏感分类算法理论 .....            | 304 |
| 13.2.2 基于 GEP 的代价敏感分类算法 .....      | 305 |
| 13.2.3 实验与分析 .....                 | 308 |
| 13.3 GEP 在特征自动抽取中的应用 .....         | 309 |
| 13.3.1 基本概念和定理 .....               | 309 |
| 13.3.2 基于 GEP 的特征集自动选取 .....       | 312 |
| 13.3.3 特征的自动聚类和特征集的自动选择算法 .....    | 315 |
| 13.4 基于 GEP 的文本分类技术 .....          | 316 |
| 13.4.1 文本的向量表示 .....               | 316 |
| 13.4.2 基于 GEP 的文本分类方法 .....        | 317 |
| 13.4.3 实验与分析 .....                 | 319 |
| 参考文献 .....                         | 319 |
| <b>第 14 章 GEP 在其他领域的应用</b> .....   | 322 |
| 14.1 GEP 在数据压缩中的应用 .....           | 322 |
| 14.1.1 数据流的基本概念 .....              | 322 |
| 14.1.2 基于 DFF-GEP 的多数据流压缩算法 .....  | 323 |
| 14.1.3 多数据流的重构算法 .....             | 323 |
| 14.1.4 实验与分析 .....                 | 323 |
| 14.2 GEP 在演化硬件研究中的应用 .....         | 327 |
| 14.2.1 进化硬件的基本概念 .....             | 327 |
| 14.2.2 二阶段电路进化方法 .....             | 328 |

---

|                                |     |
|--------------------------------|-----|
| 14.2.3 实验与分析 .....             | 332 |
| 参考文献 .....                     | 333 |
| 附录 .....                       | 335 |
| 附录 A 当前国际上部分 GEP 研究组织和资源 ..... | 335 |
| 附录 B 基本算法部分源码(C# 版) .....      | 336 |