

# 粒计算的发展现状与应用研究

赵小龙, 杨 燕

(西南交通大学信息科学与技术学院, 成都 610031)

**摘要:** 粒计算是处理信息和知识过程的一种新的研究方法, 覆盖了所有粒度的理论、方法、技术和工具的研究。该文综述了粒计算发展现状、粒计算的理论、粒计算模型与方法、粒计算在数据挖掘中的应用研究, 并给出了粒计算未来可能的研究方向。

**关键词:** 粒计算; fuzzy集; rough集; 数据挖掘

## Development Status and Application Reserach on Granular Computing

ZHAO Xiaolong, YANG Yan

(School of Information Science & Technology, Southwest Jiaotong University, Chengdu 610031)

**【Abstract】** Granular computing is an emerging research method to deal with information and knowledge processing which is an umbrella term to cover any theories, methodologies, techniques and tools that make use of granules in problem solving. This paper reviews the research of status of granular computing, the models and approaches of granular computing, the application in data mining of granular computing and future research issues are also given.

**【Key words】** granular computing; fuzzy set; rough set; data mining

1997年, 美国San Jose State大学教授T.Y.Lin第1次在他的论文中提出了粒计算<sup>[1]</sup>(granular computing, GRC), 标志着涉及多学科的一个应用研究领域产生。此后, 国外的诸多学者对它进行了研究, 提出了许多有关粒计算的理论、方法和模型, 现以成为研究模糊的、不精确的、不完整的及海量信息处理的重要工具。

### 1 粒计算理论发展现状

1979年, 美国控制论专家L.A.Zadeh第1个介绍了信息粒化(information granulation)的概念, 并认为它在Fuzzy集中有着潜在的应用; 1982年, 美国Stanford大学教授Hobbs在第9届国际人工智能大会上提出了粒度(granularity)理论, 该理论抓住了粒计算的基本特征, 以不同的粒度来概化世界, 以粒度之间的交换来处理问题; 1996年, Zadeh提出了一个新的研究分支——词计算(CW), 其目的: 为将来的智能计算以及基于词的信息系统实现计算建立一个理论基础, 标志着模糊粒度化理论的诞生; 1997年, Zadeh重新阐述了信息粒化的重要性<sup>[2]</sup>, Zadeh的工作激发了学者对粒度计算的研究兴趣。粒计算和词计算的重要性在于它能达到高档机器智商MIQ(machine intelligence quotient)的目标, 就是通过重新标记人的能力实现不用任何测量和任何复杂计算就能达到的目标。

T.Y.Lin在文献[3]中, 提出了使用邻域来作为粒计算的表达, 它是基于Rough集划分理论的一个扩展; 加拿大Regina大学教授Y.Y.Yao在文献[4]中, 讨论了粒计算的基本问题, 阐述了粒的建构和粒的计算。随后, 粒计算的研究迅速发展, 发表的文献快速增长, 目前, 国际上已形成专门的研究群体。

国内对粒计算的研究刚刚开始。张钹院士和张铃教授于2003年把原先粒度世界模型理论推广到模糊商空间理论——

模糊粒度计算方法<sup>[5]</sup>, 给出了模糊商空间的结构, 证明了利用模糊等价关系可以将原来的商空间理论推广成模糊商空间理论。刘清教授重点阐述了粒及粒计算

在逻辑推理中的应用, 讨论了信息粒的结构及其实例, 基于Rough集方法定义了决策规则粒, 构造了决策规则粒库, 定义了粒语言, 构造出了一种逻辑推理的新模型<sup>[6]</sup>。

### 2 粒计算理论的研究

#### 2.1 粒计算中最基本问题

粒计算中最基本概念是粒、粒化和粒度。粒是指一些个体(元素、点等)通过不分明关系(indistinguishability)、相似关系(similarity)、邻近关系(proximity)或功能关系(functionality)等所形成的块<sup>[2]</sup>; 粒化就是对粒进行操作; 粒度本来是一个物理学概念范畴, 指微粒大小的平均度量, 粒计算中指信息粗细的平均度量, 信息粒度是指对信息和知识细化的不同层次的度量, 从数学的角度给出粒度的定义。

**定义** 设论域U和U上的一个关系

$$R: U \rightarrow P(U) \Rightarrow U = \bigcup_{i \in X} G_i$$

则称每一个 $G_i$ 为一个粒子 $\{G_i\}_{i \in X}$ 是论域的一种粒度。其中,  $P(U)$ 表示论域U的幂集; R可代表不可区分关系、功能相近关系、等价关系、相似关系、模糊关系和一般的函数等。对 $\forall i, j \in X, i \neq j$  时有 $G_i \cap G_j = \emptyset$ , 则称 $\{G_i\}_{i \in X}$ 是论域的一种划分, 记为 $\{G_i\}_{i \in X} = [U]$ ; 当 $\forall i, j \in X, i \neq j$  时有 $G_i \cap G_j$

**基金项目:** 四川省重大基础研究基金资助项目(04JY029-001-4); 西南交通大学科技发展基金资助项目(A2004015)

**作者简介:** 赵小龙(1974-), 男, 硕士研究生, 主研方向: 粒计算, 数据挖掘; 杨燕, 副教授

**收稿日期:** 2006-07-23 **E-mail:** zxl\_jhx@126.com

$\neq \emptyset$ , 则称  $\{G_i\}_{i \in X}$  是论域的一种覆盖, 记为  $\{G_i\}_{i \in X} = \langle U \rangle$ 。

粒计算研究一般包括 2 个方面:(1)如何构造粒;(2)利用构造好的粒对问题进行计算、分析等。第(1)个方面主要处理粒的形成、表示、和语义解释等问题;第(2)个方面主要是利用构造好的粒来处理问题、解决问题。

## 2.2 粒化和粒、粒与粒、粒与层之间关系

粒和粒化、层以及粒之间的关系是粒计算研究中的关键问题, 粒在粒计算中的定义是: 论域的子集、类、对象或者聚类。粒化最初的意思是: 它是一种动作或者是一种处理, 将物体分解成粒。Zadeh 采纳了这种思想, 将一个论域划分成多粒。粒化就是将整体分解成部分, 相反, 组织就是将部分整合成整体, 基于这种定义, 粒计算的两种操作包括: 粒化和组织。

粒之间的关系可以分成两类: 相互关系和内部关系。不管那种关系, 粒化是处理粒之间的关系, 分解是将一个大粒细化, 建构是将一组相似的、不分明的、功能相近的粒组成一个较大粒。粒和粒之间关系主要有: 内部, 提炼和粗的, 划分和覆盖, 局部顺序, 相似和模糊等关系。对粒之间的关系研究得越深入, 越能促进问题的解决。

粒位于一个特定的层, 在层中粒是研究的主要对象, 不同的层对应于不同的粒类型, 粒的特点聚集了一个描述层和理解层的主要特征。层通过一个局部顺序被连接在一起, 不同层的粒是彼此相关的。

## 3 粒计算模型与方法的研究

### 3.1 Fuzzy 集理论的粒计算模型

Zadeh 认为人类在思考、判断和推理主要用语言来进行, 语言是一个很粗的粒度, 如何利用语言来进行判断和推理, 就要用到词计算。1996 年, Zadeh 提出了一个基于 Fuzzy 集理论的粒计算通用框架, 该框架基于通用约束概念来定义和建构粒, 根据模糊图或 IF-THEN 规则来表示粒之间的关系, 这种关联计算方法就是词计算。用模糊词来作为一个基本粒, 对模糊词进行推理判断, 也就是对粒进行推理判断。基于词计算的推理、决策是最贴近人类的思维形式, 而实现词计算的主要方法就是粒计算。Zadeh 进一步指出粒计算是软计算中的一个重要研究方向。

### 3.2 Rough 集理论的粒计算模型

波兰学者 Z.Pawlak 提出的粗糙集理论以某种等价关系对对象域进行划分, 给定论域  $U$ , 设  $R \subseteq U \times U$ ,  $R$  是论域  $U$  上的一个等价关系, 则  $(U, R)$  是一个近似空间, 等价关系  $R$  将论域  $U$  划分成两两互不相交的子集, 这些子集称为等价类, 等价类作为商集  $Q=U/R$  的一个元素。每个等价类可以作为一个粒, 信息表则提供了等价关系的一个具体语义解释。一个  $\forall X \subseteq U$ , 可能不是一些等价类的并, 这指出使用等价类  $R$  可能不能精确描述  $X$ , 这样可以通过 Rough 集的上近似和下近似来描述。

$$\text{arp}(X) = \bigcup_{[X]_R \subseteq X} [X]_R \quad \overline{\text{arp}}(X) = \bigcup_{[X]_R \cap X \neq \emptyset} [X]_R$$

$[X]_R = \{y | xRy\}$  是包含  $X$  的一个等价类, 低近似  $\text{arp}(X)$  是所有等价类的并, 该等价类是  $X$  的子集。上近似  $\overline{\text{arp}}(X)$  是所有等价类的并, 该等价类与  $X$  的交集不为空。

### 3.3 集合理论的粒计算模型

Y.Y.Yao 提出了基于集合理论的粒计算模型<sup>[4]</sup>, 每个粒

表示某种概念, 粒的每个元素是这个概念的实例, 粒通过信息表来构建, 用粗集的方法来处理。该模型的出发点是用集合表示粒, 粒的计算通过粒内部元素之间的运算来表示, 粒有以下几种表示, 用一个普通的集合表示; 用一个连续实数区间表示, 实数区间可以作为真值区间推理的一个基础, 如区间模糊推理、区间概率推理和粒概率推理等。它很容易扩展到研究具有模糊数的模糊数学; 用集合区间表示, 集合区间是指一个区间的两端是集合, 左端的集合包含于右端的集合, 区间集合的操作结果仍是区间集。

### 3.4 商空间理论的粒计算模型

张钊、张铃教授提出的商空间理论<sup>[5]</sup>, 建立了基于商空间理论的粒计算模型, 该模型用一个三元组  $(X, F, T)$  来描述一个问题,  $X$  表示问题的论域,  $F(\cdot)$  是一个映射, 表示论域的属性函数, 用  $F: X \rightarrow Y$  表示,  $Y$  是  $N$  维空间也可以是一般的空间,  $T$  是论域的结构, 指论域  $X$  中各元素的相互关系。分析或求解问题是指对论域  $X$  及有关的结构、属性进行分析和研究。该模型中, 当  $X$  很复杂的时, 就用比较粗的粒度来考察问题, 也就是在论域  $X$  上给出一个等价关系  $R$ , 得到一个对应于  $R$  的商集  $[X]$ , 将对应的三元组变为  $([X], [F], [T])$ , 称为对应于  $R$  的商空间, 从而将问题  $(X, F, T)$  转化为新层次的问题  $([X], [F], [T])$ , 逐步细化, 从而将问题表示成不同的粒度世界, 达到简化问题、解决问题的目的。商空间法是将不同的粒度世界与数学上的商集概念统一起来, 表示对象模型的方法, 该模型着重研究不同粒度世界之间的互相转换、互相依存的关系。

## 4 粒计算在数据挖掘中应用研究

数学上, 每个个体与某种块族结合在一起便构成了一个邻域系统, 通过二进制关系来定义邻域系统, 称为二进制邻域系统(BNS)。基于粒计算基础上的二进制邻域系统在数据挖掘中有着广泛的应用。

1998 年, T.Y.Lin 提出了二进制粒结构和二进制邻域系统的表示<sup>[3]</sup>。一个粒结构包括 4 个部分  $(V, U, B, C)$ ,  $V$  是对象空间,  $U$  是数据空间 ( $V$  和  $U$  可以为同集),  $B$  是一个精确或者是模糊邻域系统,  $C$  是概念空间, 如果  $B$  是一个二进制邻域系统, 那么  $(V, U, B, C)$  被称为一个二进制粒结构; Rough 集理论中有两个最重要的结论: 等价关系的表示和信息表的处理。但是现实世界粒化, 经常不能通过等价关系来表达, 因此, 把一般等价关系理论(Rough 集理论)转化为一般二进制关系(二进制邻域系统)是很重要的, 该文中, 用表的格式来表示二进制关系, 这个表将被称为信息表的扩展, 这样, 信息表处理将扩展为二进制关系粒结构的处理。

1999 年, T.Y.Lin 指出数据挖掘中用到的关系数据库, 总是假设论域  $U$  是一个集合, 那么对象(实体)之间是独立的, 但在现实生活中, 实体经常相关的, 作者提出了一个新的数据挖掘理论, 在原有的关系数据库中, 附加了粒计算的概念和二进制邻域系统, 用粒计算来处理二进制关系, 一系列二进制关系就是粒结构, 数据挖掘就是粒结构的处理。该理论为基于粒计算基础上的数据挖掘提供了导向。2001 年, Y.Y.Yao 提出了一个基于粒计算规则挖掘的一个框架。在 Tarski 类型中, 通过模型和可满足性的概念的基础上被开发。提出的框架是基于粒计算模型的概念的外延来定义的, 通过内涵和外延的一部分作为特征, 信息表被使用定义精确的内涵和外延, 语言的公式来定义内涵, 论域对象的子集来定义

外延,随着这个模型被提出,数据挖掘一些存在的方法可以被比较和分析。T.Y.Lin 进一步提出:逻辑上,通过一个 Cantor 集用关系结构就能建模现实世界,该论文把关系结构局限于简单种类,即二进制关系。用二进制邻域关系空间(BNS-Space)来作为最简单的建构现实模型,把二进制邻域系统结构融入粒计算应用于数据挖掘,它是基于 Cantor 集上经典关系模型的扩展。

粒计算在数据挖掘中得到了大量的具体应用。T.R.Qiu 等讨论了基于信息粒和粒计算的基础上关联规则的挖掘,以 Apriori 算法为基础,提出了基于粒计算一种新的关联规则抽取算法,和经典的 Apriori 算法相比,解决了 Apriori 算法在寻找频繁项目集时,重复扫描信息表的瓶颈问题。Y.C.Tang 等建立了以粒支撑向量机模型,能快速和精确地为生物医学数据进行二进制分类等。

现在,有关粒计算在数据挖掘中的应用文献不断出现,充分显示了粒计算作为一门理论、工具、技术在数据挖掘中的应用潜力。

### 5 粒计算的进一步工作

近年来,粒计算及其应用这一研究领域引起了国际上许多著名学者的关注,并做了许多相关的研究工作,但在粒的计算方面,所提出的一些推理方法在理论上还缺乏完备性,对实际应用还有很大的约束,同时也缺少高效的粒计算算法。粒计算覆盖了诸多领域,如人工智能、区间计算、量化理论、Rough 集理论、DS 理论、机器学习、数据挖掘等,这些理论能提供了一些具体的方法和工具解决某一个特定类型的问

题,但它们往往分散在各个领域,之间具有很少的联系,粒计算设法抽取这些领域的共性来建立一系列通用应用原理,综合它们的结果为一个整体,在统一的框架下进行研究。目前迫切需要一个粒计算的统一框架,开发出一个解决问题的通用理论。粒计算是处理不精确和部分为真值的有效方法,是建立在数学基础上的良定义理论,可以预见未来,随着信息科学的快速发展,粒计算必将发挥着越来越重要的作用。

### 参考文献

- 1 Lin T Y. Granular Computing, Announcement of the BISC Special Interest Group on Granular Computing[Z]. 1997. <http://www.cs.uregina.ca/~yyao/GrC/>.
- 2 Zadeh L A. Toward a Theory of Fuzzy Information Granulation and Its Centrality in Human Reasoning and Fuzzy Logic[J]. Fuzzy Sets and Systems, 1997, 90(2): 111-127.
- 3 Lin T Y. Granular Computing on Binary Relations I: Data Mining and Neighborhood Systems, II: Rough Set Representations and Belief Functions, Rough Sets in Knowledge Discovery[M]. Physica-Verlag, 1998: 107-140.
- 4 Yao Y Y. Granular Computing: Basic Issues and Possible Solutions[C]//Proceedings of the 5<sup>th</sup> Joint Conference on Information Sciences. 2000: 186-189.
- 5 张 铃, 张 钺. 模糊商空间理论[J]. 软件学报, 2003, 14(4): 770-776.
- 6 刘 清, 刘 群. 粒及粒计算在逻辑推理中的应用[J]. 计算机研究与发展, 2004, 41(4): 546-551.

(上接第 100 页)

从就绪表中取最高就绪任务的优先级汇编结果中“ldr”、“str”和其他指令分别为改动前 3、1、2;改动后为 3、1、(6、5),(6、5)表示 if 语句的两个分支。只增加 3 或 4 条一般指令。将任务置为就绪态和脱离就绪态指令虽有小的调整,但“ldr”和“str”指令数一样,其他语句相差一条“add”。所以可以推断代码修改前后的时钟消耗相差不多。

### 3.2 时钟消耗实验

为了测试 MCU 的 Fcclk 时钟消耗,实验中利用了 LPC2294 MCU 中的定时器 1,来捕获系统时钟。并且在系统的配置文件 config.h 中设置 Fpclk=Fcclk (#define Fpclk Fcclk)。然后分别在以上 3 个表格的各代码段前后,加入了测试代码。所需代码和相关注解如下:

```
OS_EXT volatile INT32U timesOfFpclk; //in ucos_ii.h, 测试
//用全局变量
//加在测试代码所在文件上,防止相关寄存器没有被预定义
#ifndef T1TC
#define T1TCR (*(volatile unsigned long *) 0xE0008004)
#define T1TC (*(volatile unsigned long *) 0xE0008008)
#define T1PR (*(volatile unsigned long *) 0xE000800C)
#endif
//测试时钟消耗的代码,加在被测试代码的前面
T1TC=0;//测试前置时钟计数器为 0
T1PR=0;//时钟不分频
T1TCR=0x01;//开始计数时钟周期
...被测试代码
//测试时钟消耗的代码,加在被测试代码的后面
```

```
timesOfFpclk=T1TC;//将时钟计数器的值赋值给全局变量
T1TCR=0x00;//停止计数时钟周期
试验后得到数据如表 7 所示。
```

表 7 试验数据 (Fcclk)

被测代码段	扩充前	扩充后	多消耗的时钟
得到最高优先级数	106	114,115	8,9
将任务置为就绪态	285	285	0
将任务脱离就绪态	282	282	0

所有的结果和前文的预期完全吻合,说明该方案具有极高的性价比。

### 4 结论

通过以上的推理和试验,可以认为该方案是一个很好的任务扩充解决方案。它充分地利用了 ARM 核的相关特征,最大程度地继承 uC/OS-II 系统优势。其完全扩充版反复的运用表明,系统比较稳定可靠,可以在实际项目中运用。

### 参考文献

- 1 Andrew N S, Dominic S, Chris W. ARM System Developer's Guide—designing and Optimizing System Software[M]. 北京:北京航空航天大学出版社, 2005.
- 2 uC/OS-II for ARM Processors[Z]. 2006. <http://www.Micrium.com>.
- 3 ARM Developer Suite Version I.2 CodeWarrior IDE Guide[Z]. 2006. <http://www.arm.com/>.
- 4 Jean J L. uC/OS-II The Real-Time Kernel[M]. 2<sup>nd</sup> ed. 北京:北京航空航天大学出版社, 2003.