

粒度计算的三种主要方法

程伟^{1,2}, 石扬¹, 张燕平¹

(1. 安徽大学 计算智能与信号处理教育部重点实验室, 安徽 合肥 230039;
2. 安徽电子信息职业技术学院, 安徽 蚌埠 233040)

摘要: 粒度计算(Granular Computing, GrC)是新近兴起的人工智能研究领域的一个方向,它覆盖了所有有关粒度的理论、方法、技术和工具的研究。它是词计算理论、粗糙集理论、商空间理论、区间计算等的超集。词计算理论、粗糙集理论、商空间理论是粒度计算的三种主要方法。文中着重介绍了这三种粒度计算的基本理论、模型和方法,以及它们之间的关系,并展望了进一步的研究方向。

关键词: 粒度计算;词计算理论;粗糙集理论;商空间理论

中图分类号: TP18

文献标识码: A

文章编号: 1673-629X(2007)03-0091-04

Three Primary Methods of Granular Computing

CHENG Wei^{1,2}, SHI Yang¹, ZHANG Yan-ping¹

(1. Ministry of Education Key Lab. of Intelligent Computing & Signal Processing, Anhui Univ., Hefei 230039, China;
2. Anhui Vocational College of Electronics & Information Technology, Bengbu 233040, China)

Abstract: Granular computing(GrC) is a new research field in artificial intelligence, and it covers all research relate to granular, including the theory, method, technique and skill. The theory of fuzzy information granulation, rough set theory and the theory of quotient space are three primary models of granular computing. Emphasizes the basic theories, models, approaches and the relations of them. Finally the further investigation objects are proposed.

Key words: granular computing; theory of fuzzy information granulation; rough set theory; the theory of quotient space

0 引言

人们在思考问题时,总是根据需要从不同侧面、不同角度反复对事物进行了解、分析、综合、推理,最后得出事物本质的性质和结论。人工智能研究者对人类这种能力进行了深入的研究,并建立了各种形式化的模型。而粒度计算,就是对上述问题研究的一个方面。

粒度就是求解问题时,取不同大小的对象进行研究。可将原来“粗粒度”的大对象分割为若干“细粒度”的小对象,或者把若干小对象合并成一个大的粗粒度对象,进行研究。“人类智能的公认特点,就是人们能从极不相同的粒度(Granularity)上观察和分析同一问

题。人们不仅能在不同粒度的世界上进行问题求解,而且能够很快地从一个粒度世界跳到另一个粒度的世界,往返自如,毫无困难。这种处理不同世界的能力,正是人类问题求解的强有力的表现”。

最近 Zadeh 在讨论模糊信息粒度理论^[1-3]时,认为人类认知有三个主要概念,即:粒度(granulation, 包括将全体分解为部分)、组织(organization, 包括从部分集成为全体)和因果(causation, 包括因果的关联),并在此基础上提出粒度计算。概括地说,粒度计算是以信息颗粒为核心的表示信息和计算信息的理论。而所谓信息颗粒,是人们描述问题时的一些基本集合:其内部元素一般可以用数值表示,由于彼此的相似性、功能相近、空间相邻或者一致性等而被归纳在一起。信息粒度概念是对于现实的一种抽象,其目的是为了建立一种有效的、以用户为中心的对于客观世界的认知体系,并促进人类对于客观世界乃至虚拟世界的理解。可以看出,相对于以往的数值计算,粒度计算是面向知识的计算,而前者是面向数据的计算。Zadeh 认为,粒度计算是一把大伞,它覆盖了所有有关粒度的理论、方

收稿日期:2006-06-20

基金项目:安徽省教育厅重点自然科学基金项目(2006KJ015A);安徽省自然科学基金项目(0504200208);安徽省教育厅自然科学基金项目(2005kj053)

作者简介:程伟(1964-),男,安徽阜阳人,副教授,硕士研究生,CCF 会员,研究方向为人工神经网络、机器学习、粒度计算;张燕平,教授,硕士生导师,博士,CCF 会员,研究方向为人工神经网络、机器学习、人工智能在交通、金融及农业中的应用。

方法论、技术和工具的研究。指出:“粗略地说,粒度计算是模糊信息粒度理论的超集,而粗糙集理论和区间计算是粒度数学的子集”。

按 Zadeh 粒度计算的定义, Zadeh 的“词计算理论”(Theory of Words Computing), Pawlak 的“粗糙集理论”(Theory of Rough Set), 张钹院士和张铃教授提出的“商空间理论”(Theory of Quotient Space)都属于“粒度计算”范畴,它们也是粒度计算的三种主要的模型和方法。

1 粒度的基本问题

粒度世界主要包括信息粒度的形成、表示、粗细、语义解释;信息粒子的大小;信息粒度粗细与求解有效度的关系;信息粒度的运算法则;信息粒度之间及其与外部环境的关系等。因此粒度计算的基本问题主要有两方面:一方面是如何构建信息粒度,另一方面是如何利用粒度去计算。前者处理粒度的形成、粗细、表示和语义解释,而后者处理怎样利用粒度去求解问题。粒度世界是否构造得合理极大地影响着问题求解的效率^[4]。

定义 1(粒度) 设给定论域 X 和 X 上的一个关系 $R: X \rightarrow p(X)$, 且 $X = \bigcup_{i \in \tau} u_i$, 则称每一个 u_i 为一个信息粒子, $\{u_i\}_{i \in \tau}$ 是论域的一种粒度。其中, $p(X)$ 表示论域的幂集; R 可代表等价关系、不可区分关系、功能相近关系、相似关系、相等关系、约束、相容关系、复合关系、模糊关系、属性、投影、结构关系和一般的函数等。

当 $\forall i, j \in \tau, i \neq j$, 且 $u_i \cap u_j = \emptyset$, 则称 $\{u_i\}_{i \in \tau}$ 是论域的无重叠粒度划分, 简记为 $\{u_i\}_{i \in \tau} = [X]$; 当 $\exists i, j \in \tau, i \neq j$, 且 $u_i \cap u_j \neq \emptyset$, 则称 $\{u_i\}_{i \in \tau}$ 是论域的一种覆盖, 简记为 $\{u_i\}_{i \in \tau} = \langle X \rangle$ 。

定义 2(粒子的大小) 设 X 是给定的一个论域, 粒度划分 $X = \bigcup_{i \in \tau} u_i$, 则称粒子 u 的大小为 $d(u) = \text{Card}(u) = |u| = \int_u dx$ 。注意: 当论域为离散情形时, 积分表示信息粒子 u 所含个体的总个数, 也可能是可列个; 当论域为连续状态时, 积分表示信息粒子 u 长度的度量值, 也可能是无穷大或不可数; 当 u 为模糊信息粒子时, 公式中的 u 表示集合 $\{x \mid \mu_u(x) > 0, \forall x \in X\}$, $\mu_u(x)$ 为 u 的隶属函数。

定义 3(信息粒度粗细) 设 \mathfrak{R} 是论域上关系的全体, 且 $R_1, R_2 \in \mathfrak{R}$, 若对 $\forall x, y \in X, xR_1y \Rightarrow xR_2y$, 则称 R_1 比 R_2 细, 简记为 $R_2 < R_1$ 。一个关系代表一种分类, 因此, 也可表示粒度粗细。设 $R_0 < R_1 < R_2 < \dots < R_{\text{end}}$, 表示一个嵌套关系簇, 其中 R_0 代表论域本身是一个等价类, 即最粗的划分; R_{end} 代表 $\forall x, y \in X$,

$xR_{\text{end}}y \Leftrightarrow x = y$, 即最细的划分; 其他的表示中间层次的划分。注意: 同一论域的粒度之间存在不能比较粗细的情形。

众所周知, 信息粒度的粗细影响着计算复杂度和问题的求解效率。在问题求解过程中, 同一个粒度世界或不同粒度世界所要求描述的信息含量和相互变换决定了信息粒度的粗细优化。粒度计算的目的就是在误差允许的范围内, 尽量找到计算复杂度最小的足够满意的可行近似解。因此, 可以认为粒度计算是降低计算复杂度的有效工具。总之, 如何在问题求解时选择恰当的粒度层次, 以使求解效率达到最佳, 这是粒度计算的一个关键内容^[4]。

2 粒度计算的三个主要的模型和方法

2.1 基于模糊集合论的词计算理论

集合论是现代数学的基础。模糊集合论是近三十年来发展起来的一门新的数学理论, 亦是一项新的数学工具, 是继经典数学、统计数学之后的又一新发展。美国数学家、控制论专家 L. A. Zadeh 于 1965 年在 Information and Control 杂志上发表著名论文 Fuzzy Sets, 提出模糊集概念, 奠定了模糊集合论基础。在此基础上, 于 1979 年首次提出并讨论了模糊信息粒度化问题, 推动了模糊逻辑理论及其应用的发展, 但在当时未引起足够的重视。接着, Zadeh 在 1996 年提出“词计算理论”^[1], 标志着模糊粒度化理论的诞生。其旨在解决利用自然语言, 进行模糊推理和判断, 以实现模糊智能控制。Zadeh 认为人类是用语言进行各种思考和推理的, 不同的词就表示不同的粒度, 那么如何表示它们呢? 一般来说要用“语言”、“词(word)”来表示概念, 这牵涉到“词计算”问题。Zadeh 提出的词计算理论模型^[2]: 标准化形式, If X is R , then Y is B 。描述了论域 U 上的语言粒度 X 被 R 约束, 语言粒度 Y 被 B 约束。一般化的形式, X is $r R$; 其中, is r 表明了 R 对 X 的约束方式, r 是二个离散变量, 其值决定了 R 对 X 的作用。主要的模糊约束传播规则:

$$\frac{(f(x_1, x_2, \dots, x_n) \text{ is } A)}{(q(x_1, x_2, \dots, x_n) \text{ is } q(f^{-1}(A)))}$$

其中分子表示初始数集中引入的约束, 分母表示对问题 $q(x_1, x_2, \dots, x_n)$ 的约束, $f^{-1}(A)$ 表示模糊关系 A 在映射 $f: U \rightarrow V$ 下的预像。通过约束来定义模糊粒度, $G = \{X \mid X \text{ is } r R\}$, 由约束的不同类别, 可以获得各种各样粒度。Zadeh 的模型奠定了词计算理论的基础, 它能将模糊 If - then 规则形式化地表述为: If X is $r_1 A$, then Y is $r_2 B$, 这里的 r_1 和 r_2 可以代表不同的约束种类或同一种约束。沿 Zadeh 的模糊集论的方

向,用模糊数学的方法进行有关粒度计算的方法和理论的研究,就构成“粒度计算”的一个非常重要的方法。

2.2 基于粗糙集理论的粒度计算

1982年,波兰学者 Pawlak 教授提出了用粗糙集理论研究不完整数据、不精确知识表达、学习、归纳的方法^[5]。该理论在分类意义下定义了模糊性和不确定性的概念,是一种处理不确定、不相容数据和不精确问题的新型数学工具,其主要思想就是在保持分类能力不变的前提下,通过知识约简,导出问题的决策和分类规则^[6]。这套方法是与用概率统计和模糊集合论处理不确定问题完全不同的。所谓粗糙集方法,是基于一个机构(或一组机构)关于一些现实的大量数据信息,以对观察和测量所得数据进行分类的能力为基础,从中发现推理知识和分辨系统的某些特点、过程、对象等。粗糙集理论不仅为信息科学和认知科学提供了新的科学逻辑和研究方法,而且为智能信息处理提供了有效的处理技术。粗糙集理论的特点是不需要预先给定某些特征或属性的数量描述,而是直接从给定问题的描述集出发,通过不可分辨关系(等价关系)确定给定问题的近似域,从而找出该问题中的内在规律。粗糙集理论的出发点是:根据目前已有的、对给定问题的知识将问题的论域进行划分,然后对划分后的每一组成确定其对某一概念的支持程度,即:肯定支持此概念、肯定不支持此概念和可能支持此概念。在粗糙集理论中,以上三种情况分别用三个近似集合来表示:正上限定域、下限定域和边界^[7]。

定义: U 为一对象集合,称为论域, R 为 U 上的一个由对象属性集 Ω 确定的等价关系,即 R 是 U 的一个划分,称 $\langle U, R \rangle$ 为近似空间。对于任何 $P \subseteq R$ 且 $P \neq \emptyset$, 则 $\cap P$ 也是 U 上的一等价关系,称 $\cap P$ 为 P 上不可分辨关系,记为 $\text{Ind}(P)$ 。

对于 $X \subseteq U$, 称 $R_-(X) = \bigcup \{Y_i \in U \mid \text{Ind}(R) : Y_i \subseteq X\}$ 是 X 的下近似集; $R^-(X) = \bigcup \{Y_i \in U \mid \text{Ind}(R) : Y_i \cap X \neq \emptyset\}$ 是 X 的上近似集; $R^+ - R_-$ (X) 是 X 的边界域。子集 $X \subseteq U$ 称为 U 上的一个概念,形式上,空集也视为一个概念;非空子族集 $P \subseteq R$ 所产生的不可分辨关系 $\text{Ind}(P)$ 的所有等价类关系的集合即 $U \mid \text{Ind}(P)$, 称为基本知识,相应的等价类称为基本概念;特别地,若关系 $Q \in R$, 则关系 Q 就称为初等知识,相应的等价类就称为初等概念。当 $R^+ \neq R_-$ (X) 时,就称 X 为粗糙集,从而创立了“粗糙集理论”。

目前粗糙集理论已被广泛应用于各个领域,特别是数据挖掘领域,并获得成功。

2.3 基于商空间的粒度计算

张钹院士和张铃教授在研究问题求解时,独立地

提出了商空间理论^[8]。商空间法可用一个三元组 (X, f, T) 描述一个问题。 X 表示问题的论域; $f(\cdot)$ 表示论域的属性,可用函数 $f: X \rightarrow Y$ 表示; 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 (X, f, T) , 是指对论域 X 及其有关的结构、属性进行分析、研究。对论域 X , 在其上给定一个等价关系 R , 对应与 R 的商集 $[X]$, 然后将 $[X]$ 当作新的论域, 对它进行分析、研究。故商集是将等价类看作新元素而构成的新空间, 自然得到一个较粗粒度的世界 $[X]$ 。商空间法就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, 即以商集作为不同粒度世界的数学模型的方法。问题的不同粒度表示对应于不同的等价关系 R , 也就是不同的粒度, 只不过是对论域进行不同的划分而已。因此, 划分就是构成不同粒度世界的方法。关于商空间有如下的重要结论:

(1) 商空间模型中推理的“保假原理”(或“无解保持原理”)。

设 (X, T) 是拓扑空间, $p: (X, T) \rightarrow ([X], [T])$ 是自然投影, 则 p 是连续的。若 $A \subset X$ 是连通集, 则 $p(A)$ 是 $[X]$ 中的连通集。

这一结论表明, 一个问题在原论域 X 中有解, 在适当的粗粒度论域上 $[X]$ 也有解。反之, 若粗粒度论域上无解, 则原问题必无解。这样, 在问题求解中, 可以先在一个适当粗的粒度上进行, 删去那些不连通的等价类, 然后再在剩余部分中考虑细节, 使问题得到简化。

(2) 商空间模型中推理合成的“保真原理”。

所谓“保真原理”, 是指若命题在两个较粗粒度的商空间中是真的, 则(在一定条件下), 在其合成的商空间中对应的问题也是真的。

设在两个较粗空间 X_1, X_2 上进行求解, 得出对应的问题有解, 在其合成的空间 X_3 上问题也有解。利用“保真原理”也可达到降低求解的复杂性目的。设 X_1, X_2 的规模分别为 s_1, s_2 , 因为一般情况下, X_3 的规模最大可达到 $s_1 s_2$ 。于是将原来要求解规模为 $s_1 s_2$ 空间中的问题, 化成求解规模分别为 s_1, s_2 的两个空间中的问题。即将复杂性从“相乘”降为“相加”。

3 三种粒度计算之间的关系

商空间理论、粗糙集理论、词计算理论三者都是描述人类能按不同粒度来处理事物的能力的模型。

商空间理论、粗糙集理论的共同之处在于: 认为概念可以用子集来表示, 不同粒度的概念可以用不同大小的子集来表示, 所有这些表示可以用等价关系来描述。但讨论的着重点有所不同, 商空间理论着重点是研究不同粒度世界之间的互相转换、互相依存的关系,

是描述空间关系学说的理论;而粗糙集理论主要是研究粒度的表示,刻画和粒度与概念之间的依存关系。更主要的不同在于:商空间理论是在论域元素之间存在有拓扑关系的情况下进行研究的,即论域是一个拓扑空间,而现在的粗糙集理论其论域只是简单的点集,元素之间没有拓扑关系。从这个角度说,粗糙集只是商空间理论的一个特例,即不同粒度均在同一种空间结构中进行,没有粒度空间的变化^[9]。另外,粗糙集是在给定的知识基上求解对应的问题,如求集合的 R^- 上近似和 R_- 下近似,我们是在 (X, T) 中讨论各商空间之间的关系,求相应的(各种意义下)上近似空间和下近似空间。从这个角度看,可以说粗糙集是微观的粒度计算,商空间理论是宏观的粒度计算。这两个理论都是建立在等价关系之上,所以可以将两者结合起来。

与商空间理论相比,词计算理论主要是讨论粒度的表示问题,即当人类进行各种思考和推理时,都离不开粒度,于是用“语言”、“词(word)”来表示,进而牵涉到“词计算”问题的“模糊数学”方法,而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,故两者具有等价性^[10]。

4 结束语

以上简单介绍了商空间理论、词计算理论、粗糙集等粒度计算方法之间的关系。可以看出这三个不同的粒度计算理论,从思考问题的出发点和解决问题的任

务,都不尽相同,各有千秋。但是三者都有一个共同的特点,那就是都考虑到人类智能中,有从不同粒度思考问题的这一特点。相信将商空间理论与粗糙集理论以及模糊数学方法相结合,取长补短,将能给出一个更强有力的粒度计算理论和方法。

参考文献:

- [1] Zadeh L A. Fuzzy logic = computing with words[J]. IEEE Transactions on Fuzzy Systems, 1996, 4: 103 - 111.
- [2] Zadeh L A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic[J]. Fuzzy Sets and Systems, 1997, 19: 111 - 127.
- [3] Zadeh L A. Announcement of GrC[EB/OL]. 1997. <http://www.cs.uregina.ca/yyao/GrC/>.
- [4] 李道国, 苗夺谦, 张红云. 粒度计算的理论、模型与方法[J]. 复旦学报, 2004(5): 838 - 841.
- [5] Pawlak Z. Rough Sets[J]. International Journal of Computer and Information Science, 1982, 11: 341 - 356.
- [6] 张媛, 张铃, 张燕平. 粗糙集算法及其应用[J]. 微机发展, 2005, 15(4): 17 - 18.
- [7] 许中卫, 李龙澍. 基于粗糙集理论的数据挖掘算法研究[J]. 微机发展, 2001, 11(1): 6 - 9.
- [8] 张钊, 张铃. 问题求解的理论及应用[M]. 北京: 清华大学出版社, 1990.
- [9] 张燕平, 张铃, 夏莹. 商空间理论与粗糙集的比较[J]. 微机发展, 2004, 14(10): 21 - 24.
- [10] 张铃, 张钊. 模糊商空间理论(模糊粒度计算方法)[J]. 软件学报, 2003, 14(4): 770 - 776.

(上接第 90 页)

该主键在缓存中查找,如果存在对应记录,则直接返回该对象;如果不存在,则在数据库中进行查找,并将查找结果存放于缓存中,然后将查找结果返回给应用程序。在实际应用中,对于不同的对象可能需要不同的缓存形式,因此,本系统将缓存实现为一个可配置的形式,在使用时可以根据需要灵活地选择不同的缓存。

(3) 事务管理器。

事务管理器为持久层提供事务管理的功能,持久层实例将从事务管理器中取得数据库连接。本系统并不打算独立实现事务管理,而是将它委托给底层事务管理实体,如 JDBC 事务或者 JTA 事务。

4 结束语

文中对对象关系映射进行了简要的介绍,并分析了半自动对象关系映射和全自动对象关系映射的特点,给出了半自动对象关系映射的一个设计。该持久层具有一定的对象关系映射功能,使应用程序开发人

员能够以一种统一的、面向对象的方法进行对象的存取,降低开发负担,同时有效地分离了业务逻辑与数据库访问逻辑,有利于应用系统的可维护性和可扩展性。

参考文献:

- [1] 孙卫琴. 精通 Hibernate: Java 对象持久化技术详解[M]. 北京: 电子工业出版社, 2005.
- [2] Ambler S W. The Design of a Robust Persistence Layer for Relational Databases [EB/OL]. 2005. <http://www.amblysoft.com/downloads/persistenceLayer>.
- [3] 夏晰, 曹晓钢, 唐勇. 深入浅出 Hibernate[M]. 北京: 电子工业出版社, 2005: 468 - 469.
- [4] Ambler S W. Mapping objects to relational databases What you need to know and why[EB/OL]. 2000. <http://www.ibm.com/developerworks/webservices/library/ws-mapping-to-rdb/>.
- [5] Ambler S W. Mapping Objects to Relational Databases: O/R Mapping In Detail[EB/OL]. 2000. <http://www.agiledata.org/essays/mappingObjects.html>.