第 37 卷 第 2 期
Vol.37 No.2

计 算 机 工 程
Computer Engineering

2011 年 1 月
January 2011

• 博士论文 •

文章编号：1000—3428(2011)02—0022—03 文献标识码：A 中图分类号：TP311.52

模糊集、粗糙集和商空间理论的比较研究

赵立权

(南京财经大学电子商务省重点实验室，南京 210003)

摘 要：针对模糊集、粗糙集及商空间理论，从对智能的理解、粒的表示、粒度的定义和粒的关系这 4 个方面进行比较。分析比较结果可知，它们的共同之处有：用集合定义粒，用粒描述知识；不同之处有：词计算和粗糙集理论分别从微观角度研究词的推理和属性的约简，而商空间理论是从宏观角度研究粒度的变化规律。

关键词：粒度计算；商空间；粗糙集；模糊集

Comparison Research on Fuzzy Set, Rough Set and Quotient Space Theory

ZHAO Li-quan

(Key Lab of Electronic Business, Nanjing University of Finance and Economics, Nanjing 210003, China)

【Abstract】 The fuzzy set, rough set and quotient space theory are compared in details from the viewpoints of understanding of intelligence, representation of granules, measurement of granules and relationship among granules. Their commons are to think problem from a granular perspective, to define a granule by a set and to represent a knowledge by a granule; their main difference is that theory of computing with words and rough set theory focus on reasoning of words and attribute reduction from microcosmic level respectively, while quotient space theory discusses the translation rules of granules among the worlds with different grain-size from macrocosmic level.

【Key words】 granular computing; quotient space; rough set; fuzzy set

DOI: 10.3969/j.issn.1000-3428.2011.02.008

1 概述

粒度原本是一个物理学的概念，是指对微粒大小的平均度量。在这里被借用作为对数据信息和知识粗细的平均度量，用于从宏观或微观层面上分析和处理信息。在人类的认识活动中，粒度的思想无处不在，人们观察、度量、定义和推理的实体都是粒度。

文献[1]提出并讨论了模糊信息粒度，以元素属于给定概念(信息粒)的隶属程度作为粒度，推动了模糊逻辑理论及其应用的发展。进而提出词计算理论，认为人类认知的 3 个主要概念是粒化(granulation，包括将全体分解为部分)、组织(organization，包括从部分集成全体)和因果(causation，包括因果的关联)，人类在进行思考、判断和推理时主要是用语言进行的，而语言本身就是粒度。文献[2]提出粗糙集理论，认为“人的知识就是一种分类的能力”，这个观点可能不是很完备，但却非常精炼。它用论域中的子集来表示概念，给定了论域上的一簇子集，就相当于给定了一组知识。这样在论域中给定了一个等价关系，就给定了一个知识基，然后再讨论一个一般的概念如何用这个知识基来表示。文献[3]提出一种粒度理论，并首次提出粒度(granularity)的概念，用谓词或函数项定义了粒度和划分粒度，但没有给出粒度的计算准则，将一个表示待求解的整体问题的逻辑公式用粒度理论分成若干小问题或子公式，然后分别对这些子公式求解并最后合并成整体公式的解。文献[4]进行了关于粒度问题的讨论，并为这种“粒度世界模型”建立了一整套理论和相应的算法，将其应用于启发式搜索、路径规划等方面，取得了较大的成功。

学者们认为“人类智能的一个公认的特点，就是人们能从极不相同的粒度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的求解，而且能够很快地从一个粒度世界跳到另一个粒度世界，往返自如，毫无困难。这种处理不同粒度世界的能力，正是人类问题求解的强有力的表现”[4]。这段话生动描述了粒度计算的特征，这也是建立商空间理论的出发点。然而，粒度计算作为一个专业术语是由文献[5]提出，认为粒度计算包含所有有关粒度的理论、方法论、技术和工具。近年来，关于粒度计算的文章和专著也越来越多，粒度计算的应用领域越来越广，已成为计算智能领域研究的重要热点，对它的研究将对复杂的智能系统的设计和实现产生深远的影响。

2 粒度计算的主要理论

2.1 词计算理论

模糊集理论是经典理论的推广，它认为元素总是以一定的程度属于某个集合，也可能以不同的程度属于几个集合。它研究的是一种不确定性现象，这种不确定性是由事物之间差异的中间过渡性所引起的划分上的不确定性，是事物本身固有的，它摆脱了经典数学中的二元性(非此即彼)，使得概念的外延具有一种模糊性(亦此亦彼)。

基金项目：国家“973”计划基金资助项目(2004CB318108)；教育部博士点基金资助项目(20040357002)；江苏省教育厅重大基础研究基金资助项目(07KJA52004)；江苏省自然科学基金资助项目(BK2009349)；江苏省教育厅留学基金资助项目

作者简介：赵立权(1970－)，男，副教授、博士，主研方向：人工神经网络，人工智能，粒度计算理论及其应用

收稿日期：2010-06-10 **E-mail：**zhaoliquan2004@163.com

在模糊集理论中用隶属函数来表示模糊有一定的局限性，如表达的概念缺乏前后联系、逻辑表达和算子实现都很复杂等，如何使它们能够更符合人类的思维特点，就是词计算的目的。词计算的一般形式化模型如下：

If X isr1 A then Y isr2 B

它描述的是如果 X 被 A 约束，则 Y 被 B 约束。其中，X, Y 为论域 U 上的变量；isr 为变量系词；r 为离散变量，表示约束的类型，主要有等式约束(r=e)、可能约束(r=blank)、交约束(r=c)、并约束(r=d)、真值约束(r=v)、概率约束(r=p)等。通过这些约束来定义模糊粒：G={X|X isr R}，根据约束的不同可得到不同类型的粒，通过对不同约束的组合可以得到相应的卡氏粒。再利用自然语言对粒进行标记，从而建立了词计算的理论基础。

2.2　粗糙集理论

粗糙集理论是处理不确定性问题的一种手段，它无需借助于数据以外的先验信息就可对数据进行比较客观的处理。它是利用等价关系将集合中的元素进行分类，生成集合的某种划分，与等价关系相对应。根据等价关系的性质，同一分组(等价类)中的元素是不可分辨的，这样对信息的处理就可以在等价类上进行。

粗糙集的基本模型为(U, R)，R 为论域 U 上的一个等价关系，使用等价关系 R 将论域 U 中的元素进行分类，记为：$U/R=\{[x]_R \mid x\in U\}$。对于此问题空间中的任一概念 X($X\subseteq U$)，如果它可以写成若干个等价类的并，则称 X 是可定义的，否则称 X 是不可定义的，或粗糙的。若它是不可定义的，可以利用上、下近似集描述对象，再通过知识约简，只保留某些关键信息，从而达到简化信息的目的。

2.3　商空间理论

商空间理论将不同的粒度世界与数学上的商集概念统一起来，用一个三元组(X, f, T)，即论域、属性、结构来描述一个问题，在其论域上引入等价关系 R，对应于 R 的商集[X]，然后将[X]当作新的论域，也必有一个对应的三元组([X], [f], [T])，称为原问题空间的商空间，一个商空间就是一个信息粒。

商空间理论目的是研究不同商空间(信息粒)之间的关系、商空间的分解、合成和推理规律。它从粒度的角度出发建立了研究复杂对象的框架，通过这个框架既可以从全局出发，由粗到细、由表及里地对问题进行研究，又可以从局部出发，由低层到高层、由浅到深地对问题的特征信息及规则进行挖掘的研究，在不同的粒度世界下观察同一问题，以便得到对问题不同角度的理解，从而最终达到简化问题、解决问题的目的，这和人类思考问题的方式不谋而合。

3　3 种理论的比较

3.1　对智能的理解

三者都是从各自研究的角度阐述了对人类智能的理解，虽然表述不尽相同，但是都体现了粒度计算的基本思想。一方面，人们在求解复杂问题时，将它划分为一系列更容易管理和更小的子问题，在这些子问题上进行求解，从而降低问题求解的复杂度。另一方面，在实际应用中，由于各方面的原因，人们所获得的信息是不完全的、不确定的或者模糊的，要想完全区分不同元素是很困难的，而且有时也不需要精确解。这样人们在求解问题时只关心当前研究必需的内容，选择适当的粒度，而忽略掉那些无关紧要的细节，缩小了对问题的解的搜索范围，从而提高了问题求解的效率。

Pawlak Z 认为人类智能就是一种分类的能力，表述很简单，但也很基本。Zadeh L A 认为人类认识的 3 个主要概念是粒度、组织和因果，这和文献[3]的观点不谋而合，商空间的构造问题就相当于他提出的粒度问题，商空间的合成问题就相当于他提出的组织问题，而商空间中的推理问题就相当于他提出的因果问题。

3.2　粒的表示

三者都是将所讨论的对象的集合构成论域，都是通过子集来描述粒。商空间理论、粗糙集理论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，一个等价关系对应一个粒，它们的粒可定义为：

$$G_R=U/R=\{[x]_R \mid x\in U\}$$

只不过粗糙集主要是以 G_R 中的元素，即等价类作为研究对象。

Zadeh L A 认为概念是用“词”来表示，而描述“词”的有效方法就是模糊集理论。设 X 为论域 U 上任意一个非空符号子集，则 X 中的任一有限或无限的元素组成的字符串都对应于 X 的一个词，可见词也可以用集合来表示。其粒定义为：G={X|X isr R}，是按照约束 r 的类型进行分类的。

按照 Zadeh L A 的观点，粗糙集和商空间研究的是清晰的粒，而他研究的是模糊的粒，因为在现实中研究的对象往往是连续的，大部分情况下的粒都是模糊的，如耳和脸、头和颈、手和腕、长和短、好和坏、高和矮、胖和瘦、冷和热等。然而，粒度作为描述模糊、不确定对象的工具，它是将一子集看成是一个新的元素，就是忽视元素之间的区别，它本身就是表示模糊的手段(忽视差异就是模糊)，模糊是粒度变粗的必然产物。模糊与清晰是相对的，一个清晰的概念在比它更细的空间中就变得模糊，反之，任何一个模糊的概念也必存在一个相应的粒度空间，在其上该概念是清晰的。

例如，学生的考试成绩如果是百分制，用 0~100 的整数来表示，如果是考查课，一般将其分成 4 个粒，分别用优、良、中、差来表示。优、良、中、差相对于百分制这个粒度空间来说是模糊的，而相对于 4 个粒的这个粒度空间来说，它们又是清晰的。因此，从本质上来说，词计算理论是在一个粗粒度空间上根据专家的经验，给出这些程度副词、形容词相应的度量值，这是一种先验知识。利用这样的先验知识，在问题求解时可以降低问题求解的复杂性。

3.3　粒度的定义

词计算理论是通过词的约束来定义粒度，称$\{A_i\}$为论域 U 的一个模糊粒度，若$\{A_i\}$满足：(1) $A_i\in P(U)$ (U 的幂集)；(2) $U=\cup A_i$；(3) $\mu: A_i\to[0,1]$。其中，μ 为隶属度函数，在实际应用中经常用到以下 3 类：

1)S 函数(偏大型)

$$S(x,a,b)=\begin{cases}0 & x\le a\\ 2\left(\dfrac{x-a}{b-a}\right)^2 & a<x\le\dfrac{a+b}{2}\\ 1-2\left(\dfrac{x-b}{b-a}\right)^2 & \dfrac{a+b}{2}<x\le b\\ 1 & b<x\end{cases}$$

2)Z 函数(偏小型)

$$Z(x,a,b)=1-S(x,a,b)$$

3)Π 函数(中间型)

$$\Pi(x,a,b)=\begin{cases}S(x,b-a,b) & x\le b\\ Z(x,b,b+a) & x>b\end{cases}$$

其中 a，b 为待定的集合参数；S 函数常用来表示像老、热、高、浓等偏向大的一方的模糊现象；Z 函数常用来表示像年轻、冷、矮、淡等偏向小的一方的模糊现象；S 函数常用来表示像中年、中等、适中、平均等偏向中间的模糊现象。

粗糙集理论中的不确定性是一种基于边界的概念，每个不确定性概念都可以利用上、下近似来定义。设 R 为论域 U 上的一个等价关系，对任一 $X \in U$，定义它的上、下近似分别为 $R_(X)=\{x \mid x \in X,[x]_R \subseteq X\}$ 和 $R^-(X)=\{x \mid x \in X,[x]_R \cap X \neq \varnothing\}$。$X$ 的粗糙度定义为：$P_R(X)=1-|R_(X)|/|R^-(X)|$（$|\cdot|$表示集合的势），当 $P_R(X)=0$ 时 X 为可定义的，否则 X 为不可定义的。

模糊集理论和粗糙集理论在处理不确定性和不精确性问题方面都推广了普通的集合理论。它们都是研究信息系统中知识不完全、不确定问题的重要方法。但模糊集理论中的对象的隶属度一般是由专家直接给出，带有很强的主观性且缺乏精度的概念；而粗糙集理论中对象的隶属函数值却依赖于知识库，它可以从所需处理的数据中直接获得，所以用它来反映知识的模糊性是比较客观的。

从知识的“粒度”的描述上来看，模糊集理论中通过对象关于集合的隶属程度来近似描述的，属于同一类的不同对象间的隶属关系，它着眼于集合的模糊性，强调的是隶属程度；而粗糙集理论是通过一个集合关于某个可利用的知识库的上下近似来描述的，反映的是不同类中的对象组成的集合关系，它着眼于集合中对象间的不可分辨程度，强调的是分类。商空间理论由于是从宏观上讨论粒的变化规律，它并没有研究粒的度量方法。

3.4 粒的关系

商空间理论、粗糙集理论、词计算理论都将所讨论的对象的集合构成论域，但讨论对象之间的关系时，却各有不同。

词计算模型来源于词计算，其模型可以归结为(U, F)，其中，U 为论域；F 为论域上的属性集，它主要是研究词的表示(度量)及其推理，用模糊集理论进行词的度量，用 If-Then 规则进行推理。粗糙集模型来源于数据库，其模型为(U, F)，其中，U 为论域，F 为论域上的属性集，它是通过元素的不同属性值来描述元素之间的关系，并用元素按不同属性进行的分类来表示不同的概念粒度。商空间模型来源于分层递阶方法，其模型为(U, T, f)，其中，U 为论域；T 为论域上的结构；f 为论域上的属性集，即除了元素的属性外，还引入元素之间的关系 T(用拓扑来描述)。商空间理论研究的是有结构的模型，这在实际问题求解中是经常遇到的，如地理信息中其地理位置之间的关系就是一个距离结构；在数据仓库中各数据之间的关系可用半序来描述，它也是一种结构；在路径规划中对象所处空间的位置关系，就是一种距离的结构；在数据挖掘中的规则发现，所有的规则全体按其包含关系就构成半序结构等。当然结构在一定意义下也可以看成元素的某种属性，如优、良、中、差也有序关系，但当属性是多元的(要用多元函数来表达)时，一般不能表示为 $f(x)$，而要用 $f(x_1, x_2, \cdots, x_n)$表示，如距离要用 $d(x, y)$表示。

粗糙集理论研究的论域一般没有结构关系，只是简单的点集。词计算理论主要擅长描述由形容词、副词表达的不同粒度的概念，如非常好、很好、好、很不错、还好、大约、可能、相似、……。这些词有程度不同的差别，故在一定意义下，词计算理论也给出了描述元素之间的关系，但只限于由属性的强弱程度不同所形成的关系。

4 结束语

由以上对比可以看出，词计算理论是从微观的角度研究词的推理，粗糙集理论是从微观角度研究属性的约简，而商空间理论是从宏观角度研究粒度的变化规律。这 3 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，即都考虑到人类智能中，有从不同粒度思考问题的这一特点。如何将三者的优点结合起来，形成更强有力的粒度计算的方法和理论，是今后一个重要的研究课题。

参考文献

[1] Zadeh L A. Fuzzy Sets and Information Granularity[M]. [S. l.]: North-Holland Publishing Co., 1979.

[2] Pawlak Z. Rough Sets[J]. International Journal of Information and Computer Science, 1982, 11(5): 341-356.

[3] 张 钹, 张 铃. 问题求解理论及应用[M]. 北京: 清华大学出版社, 1990.

[4] Hobbs J R. Granularity[C]//Proceedings of the 9th International Joint Conference on Artificial Intelligence. [S. l.]: IEEE Press, 1985: 432-435.

[5] Lin T Y. Granular Computing[M]. [S. l.]: Kluwer Academic Press, 1997.

编辑 陈 文

(上接第 21 页)

信噪比，其均值最大。用 $y\sim n$ 曲线处理时，db4 小波函数的多尺度降噪的信噪比均值最大，而且每张基因芯片的信噪比均优于原始数据降噪，其降噪效果最好。

5 结束语

$y\sim n$ 曲线模型能分析处理基因芯片涵盖的大量基因表达信息，将一张零散的基因芯片数据转化为一列相互关联的信号，因此，可以利用多尺度分析进行深入研究。由降噪效果可知，$y\sim n$ 曲线比原始数据具有更好的多尺度分析效果，小波函数 db4 在 $y\sim n$ 曲线的基础上对大脑皮质的基因芯片数据进行多尺度分析效果最好。可以预见，$y\sim n$ 曲线在以后的基因芯片的数据处理方面必将有大量的应用。

参考文献

[1] 张宏怡, 张军英. 基于因果关系挖掘的概率基因调控网络的构建[J]. 计算机工程, 2007, 33(15): 26-28.

[2] Lawrence T C. Study the Gene Expression Profiles of Developing Human Fetal Central Nervous System Suing High Density cDNA Microarrays[D]. Hong Kong, China: City University of Hong Kong, 2003.

[3] 李 瑶. 基因芯片数据分析与处理[M]. 北京: 化学工业出版社, 2006.

[4] Chen Lingfeng, Zhang Xusheng, Lin Jiaming, et al. Signal-to-noise Ratio Evaluation of a CCD Camera[J]. Optics & Laser Technology, 2009, 41(5): 574-579.

[5] 罗万春, 陈 军, 伍亚舟, 等. 基于小波多尺度的人类胚胎期大脑皮层基因表达分析[J]. 重庆医学, 2009, 38(12): 1462-1463.

编辑 顾姣健