

# 不精确性:涵义与性质

黄映辉 李冠宇

(大连海事大学信息科学技术学院 大连 116026)

**摘要** 随机性和模糊性一直被认为是研究不确定性现象的两个方面。粗糙性的介入形成了解释不确定性现象的“三位一体”结构。随机性是客观属性,模糊性和粗糙性则与人的认知活动有关,故被合称为不精确性,且将之作为研究不确定性现象的主要突破方向。不精确性是人们在认知不确定性事物时必须接受的代价和缺陷;模糊性是为了获得认知的可行性和效率而放弃的对认知对象的清晰判定;粗糙性是当采用已有知识集合的两个子集近似地解释认知对象时而失去的对认知对象的细致描述。模糊性源于人们对认知对象的分类,粗糙性则与人们已有知识的不完备状态有关。不精确性的主要性质有三:主观性、基于分类、依赖已有知识。减低不精确性应从已有知识入手,即增加存量和改善有序程度。

**关键词** 不精确性,模糊性,粗糙性,不确定性

**中图法分类号** TP18 **文献标识码** A

## Impreciseness: Meaning and Property

HUANG Ying-hui LI Guan-yu

(Information Science and Technology College, Dalian Maritime University, Dalian 116026, China)

**Abstract** For a long time the randomness and the fuzziness are considered as the two aspects for explaining the uncertain phenomena, but the roughness being introduced later promotes to form the triune framework for describing the uncertainty. The randomness is an objective attribute, but the fuzziness and the roughness are related to human's cognitive activities so that are collectively known as the impreciseness. The impreciseness becomes a promising research direction for the uncertain phenomena. The impreciseness is the cost and defects which human have to bear when cognizing the uncertain objects, in detail, the fuzziness is the loss of clear understanding for a cognitive object in order to obtain the cognitive possibility and efficiency; the roughness is the lack of exact description for a cognitive object as to be explained approximatively with two subsets of the existing knowledge set. The fuzziness is derived from the human's classifying cognitive objects, and the roughness from the imperfection of human's existing knowledge. The impreciseness has three primary properties: subjectivity, classifying-based, and relying on existing knowledge. To decrease the impreciseness should start with the existing knowledge, increasing the quantity and improving the orderliness.

**Keywords** Impreciseness, Fuzziness, Roughness, Uncertainty

## 1 引言

近年来在信息科学领域对“不精确性信息”、“不精确性本体”、“不精确性推理”等的研究<sup>[1-3]</sup>日渐增多。然而,对“何谓不精确性(impreciseness)”却鲜有讨论,以致于在实际应用中常将其与“不确定性(uncertainty)”相混淆,视二者为等同关系、并列关系、包含关系的用法均有。本文尝试对此疑惑进行析解,其推进的路线为:首先回溯不确定性的诱因解释,随之分析不精确性的基础语义,进而指明不确定性与不精确性的相互关系,最后讨论不精确性的主要性质。

## 2 不确定性:客观现象及其诱因解释

### 2.1 不确定性是一种客观现象

属性只能依附于主体。不确定性说明的是客观世界中存

在着的事物,完整的说法应该是“事物的不确定性”。事物变化是客观规律,沿着时间轴线则表现为一系列状态的交替,毋庸置疑这些状态之间有着内在的联系。通常将时间在前的状态称为原因,将时间在后的状态称为结果,因果律则说明它们之间的内在联系。

若一事物从某个先前状态(原因)发展变化至某一个或少数几个后续状态(结果),则称该事物具有确定性,即符合因果律。若一事物从某个先前状态(原因)发展变化至有限多个或无穷个后续状态(结果),则称该事物具有不确定性,即背离因果律。可见,不确定性是指“事物发展结果的不确定性”。事物的发展是永恒的,不确定性就是一种客观现象。

### 2.2 随机性:最初的解释

因果律将客观世界的万千事物划分为两类:确定性事物和不确定性事物。人类对此的认知策略,一方面是避难就易

到稿日期:2009-05-20 返修日期:2009-07-22 本文受国家自然科学基金项目(60672031),辽宁省自然科学基金项目(20072142)资助。

黄映辉(1955—),男,博士,教授,CCF高级会员,主要研究方向为智能信息处理,E-mail:hyh6599@163.com;李冠宇(1964—),男,教授,CCF高级会员,主要研究方向为智能信息处理。

优先选择前者,另一方面是将后者分解或简化为前者,所对应的便是以牛顿理论为标志的确定性科学的辉煌。然而,人类一直怀有对不确定性事物的好奇:不确定性现象背后的原因到底是什么?

19世纪末发展起来的统计力学和量子力学,为人类认知不确定性现象开启了思路:造成“事物发展结果的不确定性”之现象的原因在于事物内部蕴含着的随机性(randomness),概率论的应用于是兴盛起来,其中 T. Bayes 的贡献最大。随着概率论进入大学甚至中学课堂,“不确定性就是随机性”成为根深蒂固的观念。

### 2.3 模糊性:从点状态到区间状态

随机性只能解释那些状态本身的类属关系是确定的、仅是究竟出现哪个状态是不确定的事物的发展,而对那些状态本身的类属关系就不清晰的事物则无能为力。模糊性(fuzziness)受到了关注,其开篇之作是1965年 L. Zadeh 题为 Fuzzy Sets 的论文<sup>[4]</sup>。至此,模糊性被追加纳入对不确定性现象的解释,即“不确定性=随机性+模糊性”。

随机性是客观的,模糊性则不然。二者的根本区别在于对事物发展状态的排中律(事物在某时点上非此状态即彼状态)的背离与否。事物的状态根据其形成方式可分为点状态和区间状态,前者直接在事物的发展轨迹上截取,后者则取决于人们认知事物时如何进行分类。以人的年龄为例,“35岁”是点状态,“青年、中年、老年”是区间状态。“35岁”到底是“青年”还是“中年”?只能以“模糊性”一言而蔽之。可见,模糊性对应于区间状态,与人们对事物状态的分类有关,因而具有主观随意性。若人类认知事物时能够完全基于点状态,那么模糊性也就不存在了。然而,客观事物的发展总是连续的,而人类的认知能力极为有限,将事物的连续状态人为割裂以归类为若干个区间状态,是不得已的选择。在此意义上,模糊性也是客观的(确切地说是必然的)。图1示意了随机性、模糊性与不确定性的关系。

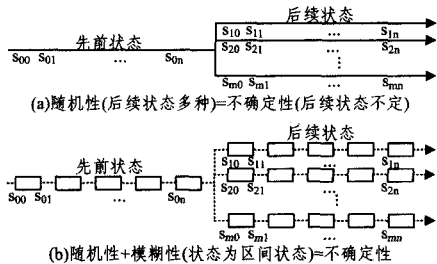


图1 不确定性的内涵:客观的随机性和主观的模糊性

如图1所示,随机性用于表示事物发展的状态之间的关系,即一个先前状态 \$s\_0\$ 有着多个可能的后续状态 \$s\_1, s\_2, \dots, s\_n\$。模糊性则是对事物发展的状态本身的描述,即先前状态和后续状态均是区间状态。因为,总是先有状态,后有状态之间的关系,而且人类为了认知的便利只能将连续的点状态归类为离散的区间状态,所以就人类考察客观事物的逻辑而言,模糊性比随机性更为基础。图1(b)示意了人类认识不确定性事物的一般图景:首先,应用分类方法将事物的发展表示为区间状态(模糊性伴随而生)的时间序列;然后,应用概率论分析对应于某个先前状态的后续状态的随机分布规律。此时的“不确定性”之概念本身是平衡的:模糊性在先、随机性在后;模糊性具有主观性,随机性具有客观性。

### 2.4 粗糙性:基于已有知识的认知

模糊性根植于客观事物差异的中介过渡性,这种差异完全是由将连续的点状态分类为跃迁的区间状态所致。经分类所得的区间状态的类型越多即越接近于点状态,区间状态本身的模糊性就越弱。可见,模糊性与如何分类有着密切的关联。迄今业已成熟的模糊性理论(包括模糊数学、模糊系统、模糊工程等)主要研究分类之后所得的模糊性对象(模糊事物、模糊概念、模糊状态等)的性质和相互关系。对人们是否能够准确、完整地认知这些模糊性对象(清晰性对象是其特例)的追问,催生了 Z. Pawlak 于1982年提出的粗糙集理论<sup>[5]</sup>。粗糙性(roughness)成为描述不确定性事物的一种新属性。

粗糙性深入分析了图1(b)中事物发展状态 \$s\_j\$ 的性质。设有论域 \$U\$, 将其表示为基本范畴 \$R\_i\$ 的集合, \$R\_i\$ 也称为等价类(equivalence class), 因为该类中的所有对象均服从同一个等价关系<sup>[6]</sup>。等价类 \$R\_i\$ 及其任一组合均是人类已有的关于论域 \$U\$ 的确定性知识。现将状态 \$s\$ 置于论域 \$U\$ 中(如图2所示)。可见,状态 \$s\$ 完整地包含了3个等价类(黑色标出),该集合称为状态 \$s\$ 的下近似集 \$\underline{R}\_s\$, 其含义是人们可用由这3个等价类构成的已有知识来理解状态 \$s\$; 状态 \$s\$ 的边界又部分地包含5个等价类(灰色标出),该集合称为状态 \$s\$ 的边界域 \$bn\_R(s)\$, 其含义是人们可用这5个等价类对应的已有知识来部分地解释状态 \$s\$; 状态 \$s\$ 的下近似集 \$\underline{R}\_s\$ 与边界域 \$bn\_R(s)\$ 两者之和称为状态 \$s\$ 的上近似集 \$\overline{R}\_s\$, 其含义是人们对状态 \$s\$ 的理解涉及到8个等价类(黑色和灰色标出)对应的已有知识。

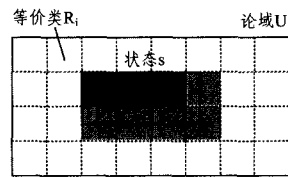


图2 状态 \$s\$ 的粗糙性:下近似集与上近似集

由图2可知,人们应用关于论域 \$U\$ 的现有知识不能“精确”地理解状态 \$s\$, 只能求助于下近似集 \$\underline{R}\_s\$ 和上近似集 \$\overline{R}\_s\$ 去“粗糙”地表述,即状态 \$s\$ 具有粗糙性。粗糙性源于下近似集和上近似集的不重合。

将上述分析应用于图1所示的模型,事物状态 \$s\$ 就是具有粗糙性的,即人们并不能利用已有的知识准确地理解它们,仅能用下近似集 \$\underline{R}\_s\$ 和上近似集 \$\overline{R}\_s\$ 予以说明。补充了粗糙性的不确定性模型如图3所示。对于状态 \$s\$ (用矩形表示), 其下近似集 \$\underline{R}\_s\$ 为矩形中的黑色区域, 上近似集 \$\overline{R}\_s\$ 为矩形的整个区域, 粗糙性的程度就对应于矩形中的白色区域(即边界域 \$bn\_R(s)\$)的大小。

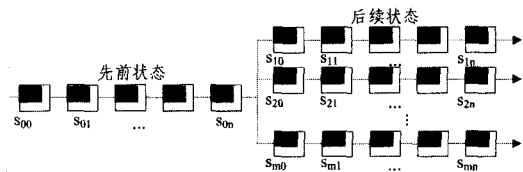


图3 不确定性的内涵:随机性+模糊性+粗糙性

本节的分析说明,不确定性是人类在认知客观事物时所遭遇的常见现象,即无法获得一个定值的、恒稳的、清晰的认知。不确定性包括随机性、模糊性、粗糙性。随机性是客观事

物自身性态和类属都是确定的,只是其能否发生无法预料,是因果律的破缺;模糊性是客观事物自身性态和类属就是不确定的,可见它是比随机性更为基本的不确定性,是排中律的破缺;粗糙性则是人类对客观事物进行分类时的先前约定(即应用怎样水平的知识库来支持认知活动),对模糊性有着直接的影响,分类的粒度大则模糊,粒度小则清晰。

### 3 不精确性:人类认知客观事物的程度

不确定性是大多数客观事物在发展过程中呈现出的“后续状态不定”的现象。人们一直在寻求这种现象背后的原因,先后提出了随机性、模糊性、粗糙性的解释。随机性是说,事物总是处在复杂系统中,其发展走向势必受到诸多外界因素的干扰,故其后续状态有多种可能。毋庸置疑,随机性是对不确定性现象最贴近的解释。客观世界本来如此,按其自身规律演变发展着。然而,当人们试图认知这种由随机性推动的不确定性现象时,问题就显现出来:受认知能力所限也是为了提高效率,人们只能对客观事物进行分类以减少认知对象的数量。简练认知对象的类型,有利于人们集中精力,同时引发了“隶属与否”的难题:某个事物在多大程度上隶属于A类或B类?模糊性由此而来,它是人们认知客观事物的“可行性”和“高效性”的代价。世上本无模糊性,完全是“人造”出来的。模糊性是说,人们为了获得认知的可行性和效率,必须对客观事物进行分类。然而,经这种分类所得的类型是否合适?这里所谓的“合适”是指,这些类型是否能被人们应用已有的知识“精确”地定义,粗糙性就是对此的度量。在大多数情况下,人们只能采用下近似集和上近似集这两个精确集来“粗糙”地理解客观事物分类后的具体类型。

不确定性现象诱发于客观事物本身固有的随机性。人们为了更好地认知不确定性现象,又增添了模糊性和粗糙性。将不确定性理解为“三位一体”,即不确定性包括随机性、模糊性和粗糙性,成为学界的主流观点<sup>[7]</sup>。对粗糙性的关注打破了图1(b)所示的不确定性的随机性与模糊性的平衡:随机性是客观属性,模糊性与粗糙性则与人们的认知活动有关,是认知客观事物程度的度量。人们对客观实在的所为是有限的,而对自身的改善却是无穷的。人们对待不确定性现象,一方面进一步完善概率理论使之更逼近随机性的真相,另一方面更注重如何恰当地分类客观事物(模糊性与认知效率之间的权衡)和如何有效地利用已有的知识资源(粗糙性与认知成本之间的权衡)。显然,后者需要更多的深入研究。因而,模糊性和粗糙性二者合称为“不精确性”,成为一个专门的研究方向。亦即,对不确定性现象的研究重点已经从随机性转移到不精确性上。这也正是“不精确性”一词近年悄然兴起的原因。

### 4 不确定性与不精确性:包含关系

哲学家们一直认为,不确定性主要有两个来源:客观事物的使然和人的认知能力所限<sup>[8,9]</sup>。本文的上述分析,将之分别归因于随机性和不精确性,且不精确性包括模糊性和粗糙性。不确定性与不精确性是因果关系,前者是现象,后者是其两大原因之一。也可将二者视为包含关系,即不确定性=随机性+不精确性。如此界定,有利于避免当今在不确定性领域的研究中常见的术语误用或混用的情况。

## 5 不精确性的主要性质

### 5.1 主观性

不精确性对应于人们为了更好地认知客观事物发展的不确定性现象(由事物内部的随机性推动)而采取的人为措施。其中,模糊性与对客观事物如何进行分类有关,粗糙性与将客观事物置于怎样的已有知识背景中有关。不精确性源于人类对客观事物的认知活动,具有明显的主观性。

有趣的是,将模糊性和粗糙性合称为不精确性,还有着语言学渊源。中文“精,择也”、“确,石头般坚固”,英文 precision 则源于拉丁语 praecedere(切断)的过去分词。可见,“精确(precision)”描述的是“处理的过程和结果”,也就是说不精确性(impreciseness)是关于人类对客观事物进行处理的程度的度量。随着对客观事物的处理的细致与深入,不精确性逐步减低,即由“粗糙”趋向“细致”,进而由“模糊”趋向“清晰”,以便于人类的认知与讨论。

### 5.2 基于分类

模糊性源于人们为了提高认知的可行性和效率而采用隶属度函数的方式对客观事物所进行的分类。该客观事物被称为论域,分类所得的各个类型被称为模糊子集,每个模糊子集被一个隶属度函数唯一地确定<sup>[10]</sup>。模糊性是指,某个具体的对象即隶属于这个模糊子集又隶属于那个模糊子集的性质。现以人为例子予以说明。人的年龄区间为0~150岁,若将同龄人作为一个集合,就有150个研究对象。若真如此,一方面认知的工作量太大,另一方面也没有必要,因为在生理、心理、性格、能力等方面21岁的人和22岁的人并无显著的差别。L. Zadeh的做法是,将人划分为“青年”和“老年”两类,并给出了那两个著名的隶属度函数<sup>[4]</sup>,相应的曲线表达如图4所示。

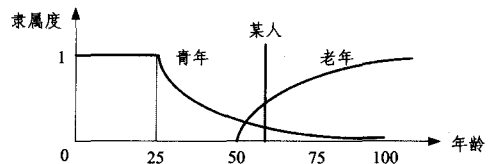


图4 Zadeh观点:每个人具有两类人的特征

这样,人类只需深入研究“青年”和“老年”两类人群的特征即可。每个人都在不同程度上具有这两类人的特征。将一个年龄确定的人置于图4中,其对应的两个隶属度函数值(被分别称为对两类人群的隶属度)表示对该人的分析和预测,例如一个60岁的人具有2%的“青年”特征和80%的“老年”特征。需要指出的是,此处的“青年”和“老年”均是Zadeh意义上的。也就是说,模糊性源于分类,模糊性的程度则取决于如何分类。

### 5.3 依赖已有知识

人们对客观事物的认知是一种比较操作,即将认知对象与储存于人脑中的已有知识进行对比。若切合,就是所谓的“认识该客观事物”。然而,完全切合和总能切合的情况毕竟很少。粗糙集理论的解决办法是,用下近似集 $R_s$ 和上近似集 $\bar{R}_s$ 来表示“粗糙”切合的情况(如第2.4节所述)。这样,粗糙性的程度就和下近似集与上近似集的差值即边界域 $bnR(s) = \bar{R}_s - R_s$ 的大小正相关,通常采用粗糙度 $r_R(s)$ 表示<sup>[11]</sup>:

$$r_R(s) = \frac{card(R_s) - card(R_s)}{card(\bar{R}_s)}$$

其中,  $card$  表示近似集的基数, 即所包含的等价类的个数。对于图 2 所示的情形,  $r_R(s) = (8-5)/8 = 0.38$ 。

论域  $U$  所包含的等价类  $R_i$  越多(即已有知识越细化), 边界域  $bn_R(s)$  的比重就会减小, 粗糙度  $r_R(s)$  也就降低了。图 5 是将图 2 的等价类增大至 4 倍的情形, 粗糙度  $r_R(s) = (21-18)/21 = 0.14$ , 较先前下降了许多。

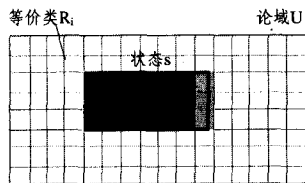


图 5 已有知识的丰富可以降低粗糙性

图 2 和图 5 的分析说明, 粗糙性并不是客观事物本身具有的属性, 而是人类对客观事物认知的“不精确性”程度。与其说是“事物的粗糙性”, 不如说是“认知的粗糙性”。粗糙性的程度与认知主体已经拥有的知识存量直接相关。例如, 同一个客观事物, 对一位科学家而言的粗糙性就要比对一位农民而言的粗糙性小得多。可见, 包含粗糙性的不精确性依赖于认知主体的已有知识, 呈现一种反向变化的关联。

**结束语** 本文将不精确性合称模糊性与粗糙性, 以便从人的认知活动和知识管理的视角对其深入考察, 进而推进对“不精确性本体”、“不精确性推理”等的研究。本文研究的主要结论如下。

(1) 不精确性源于人们对不确定性现象的认知活动。人们为了更好地认知不确定性现象, 采取了两项有力措施: 一方面是简化认知对象, 将连续发展的客观事物人为地划分为若干类型。为了继续保持连续性, 分类采用模糊子集的方式, 即模糊子集之间的边界相互交迭, 以致于某个具体事物同时隶属于多个模糊子集, 模糊性由此而来。另一方面是利用已有知识, 人类总是用已有知识来认知新事物的, 新事物之“新”就在于该事物并不能由已有知识完全解释, 可行的方案是采用与该新事物相毗邻的两个已有知识的集合(即下近似集和上近似集)来近似地说明, 粗糙性由此而来。不精确性是模糊性与粗糙性之合称。

(2) 减低不精确性的最好途径是有效的知识管理。不精确性的影响是负面的, 对其的抑制无非是从模糊性和粗糙性两方面入手。模糊性就是由于分类而导致的“亦此亦彼性”。分类越细, 越有利于判别对象隶属于何类而减小模糊性。然

而, 模糊性对应于隶属度函数, 后者又完全由主观确定, 分类越细小, 需要确定的隶属度函数就越多, 众人就更难取得一致, 使原本模糊的事情更加模糊。模糊性的减少是困难的。粗糙性的情景就要乐观得多。粗糙性就是认知对象与认知主体(人或计算机系统)的已有知识的不吻合程度, 通过有效的知识管理增加已有知识的存量, 改善已有知识的有序性, 就可减低粗糙性进而影响不精确性。

(3) 不精确性的测度尚待研究。属性能够定量表达才有应用价值。模糊性采用可信性测度, 并由此导出隶属度函数; 粗糙性采用信赖性测度, 并由此导出粗糙度函数<sup>[12]</sup>。由模糊性和粗糙性构成的不精确性应该采用怎样的测度? “不精确度函数”又应该如何定义? 对此需要进一步探讨。

## 参 考 文 献

- [1] Klinov P, Taylor J M, Mazlack L J. Interval rough mereology and description logic: an approach to formal treatment of imprecision in the Semantic Web ontologies[J]. *Web Intelligence and Agent Systems*, 2008(6): 157-174
- [2] Klinov P, Mazlack L J. Fuzzy rough approach to handling imprecision in Semantic Web ontologies[EB/OL]. <http://www.ece.uc.edu/~klinov/p/doc/conferences/2006/nafips.2006.pdf>
- [3] Klinov P, Taylor J M, Mazlack L J. Formal treatment of imprecision in the Semantic Web ontologies[EB/OL]. [http://www.ece.uc.edu/~klinov/p/doc/proposals/nsf\\_aprx\\_onto.2006.pdf](http://www.ece.uc.edu/~klinov/p/doc/proposals/nsf_aprx_onto.2006.pdf)
- [4] Zadeh L. Fuzzy sets[J]. *Information and Control*, 1965, 8(3): 338-353
- [5] Pawlak Z. Rough sets[J]. *International Journal of Computer and Information Sciences*, 1982, 11(5): 341-356
- [6] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001
- [7] 李德毅, 杜鹁. 不确定性人工智能[M]. 北京: 国防工业出版社, 2005
- [8] 柳廷廷. 科学世界图景中的不确定性[J]. *哲学研究*, 1993(5): 58-64
- [9] 鲁鹏. 论不确定性[J]. *哲学研究*, 2006(3): 3-10
- [10] 罗承忠. 模糊集引论(上册)[M]. 北京: 北京师范大学出版社, 2007
- [11] 刘清. Rough 集与 Rough 推理[M]. 北京: 科学出版社, 2001
- [12] 刘宝碇, 彭锦. 不确定性理论教程[M]. 北京: 清华大学出版社, 2005
- [13] Washington, DC: IEEE Computer Society, 2006
- [14] Bayardo R, Agrawal R. Data privacy through optimal k-anonymization[C] // *Proceedings of ICDE 2005*. Washington, DC: IEEE Computer Society, 2005: 217-228
- [11] 刘向宇, 杨晓春, 于戈. 一种基于特征类的高精度隐私保护数据发布方法[J]. *计算机科学*, 2005, 32(7增A): 368-373
- [12] 杨晓春, 刘向宇, 王斌, 等. 支持多约束的 K-匿名化方法[J]. *软件学报*, 2006, 17(5): 1222-1231
- [13] Machanavajjhala A, Gehrke J, Kifer D. l-diversity: Privacy beyond k-anonymity[C] // *Proceedings of ICDE 2006*. Washington, DC: IEEE Computer Society, 2006: 1-12
- [14] Xiao Xiaokui, Tao Yufei. Personalized Privacy Preservation [C] // *Proceedings of the ACM SIGMOD 2006*. New York: ACM, 2006: 229-240
- [15] Hyoungmin P, Kyuseok S. Approximate Algorithms for k-Anonymity[C] // *Proceedings of the ACM SIGMOD 2007*. New York: ACM, 2007: 67-78

(上接第 150 页)

- [4] Hundepool A, Willenborg L.  $\mu$ - and  $\tau$ -argus: software for statistical disclosure control[C] // *Proceedings of Third International Seminar on Statistical Confidentiality*, 1996
- [5] Sweeney L. Achieving k-anonymity privacy protection using generalization and suppression[J]. *Int'l Journal on Uncertainty, Fuzziness and Knowledge-Based Systems*, 2002, 10(5): 571-588
- [6] Meyerson A, Williams R. On the complexity of optimal k-anonymity[C] // *Proceedings of PODS 2004*. New York: ACM, 2004, 6: 223-228
- [7] Aggarwal G, Feder T, Kenthapadi K, et al. k-Anonymity: Algorithms and Hardness[R]. Stanford University, 2004
- [8] Lefvre K, DeWitt D, Ramakrishnan R. Incognito: Efficient full-domain k-anonymity[C] // *Proceedings of the ACM SIGMOD 2005*. New York: ACM, 2005: 49-60
- [9] LeFevre K, DeWitt D J, Ramakrishnan R. Mondrian Multidimensional K-Anonymity [C] // *Proceedings of ICDE 2006*.