

生物医学知识整合论（IX）

——献给电子病历

包含飞

上海中医药大学中医学信息化-标准化研究室

201203, bhf2002@online.sh.cn

摘要 本文讨论了病历陈述性数据的认知目标和 6 类广义数据生成器，病历数据的生成属性、表达属性及数据单元的基本结构、认知属性及顺逆向认知方向，数据的语义运算，最后应用实例再次讨论了生物医学知识整合论的知识背景空间的实质及与推理的关系并描述了建立知识背景空间的过程。

关键词 生物医学知识整合论 电子病历 数据属性 知识背景空间

（ I ） 病历陈述性数据的认知目标和认知器

本文为生物医学知识整合论（The Theory of BioMedical Knowledge Integration, BMKI）^[1-12]专为开发电子病历数据语义的基础理论研究。讨论仅限于讨论病历的陈述性或观测性数据的属性。

一般来说，病历的陈述性数据（declarative data）的认知目标：（1）是否存在异常事物？（2）如果存在，则在已知疾病集合中指定为何种异常？（3）如果存在，但在已知疾病集合中无法指定，则是否能确定发现了新的异常。

在人类的认知过程中,医护人员作为认知主体与认知对象（病人、疾病、症状、病因、病机等）的关系如图 1。而病历为医生与护士对疾病和病人的认知和处理过程的记录，反映了二者之间的各种关系。深刻分析和认识病例数据的信息学属性，是我们开发强有力的电子病历的第一步。

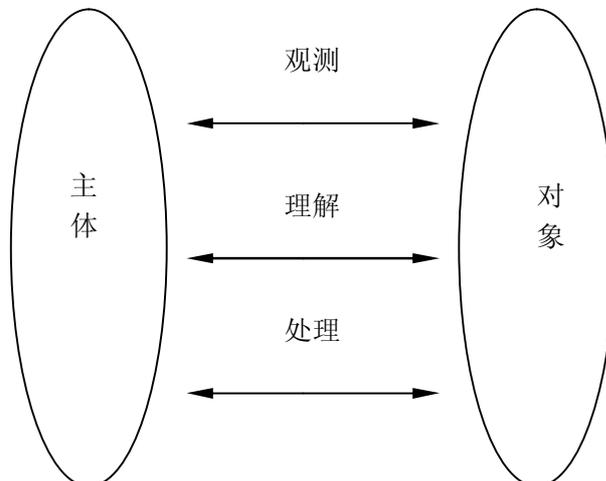


图 1 人类作为认知主体与认知对象的关系

认知主体利用广义的认知器（观测器/判断器）实现认知过程。广义的认知器又有客观认知器和主观认知器之分。客观认知器指人们设计制造的设备、仪器或试剂等；主观认知器指人的固有的认知功能器，可进一步分为感官层次认知器，一般心理层次认知器和科学（专业）知识层次认知器三类。认知对象的有关内容也有表浅和深潜之分：可观测部分，一般机制（一般可理解部分），深层机制（专业可理解部分）。一般机制和深层机制往往需要各种心理或意识层次的功能参与，以获得有意义的的数据。粗略地讲，主观认知器的3个层次与认知对象的3层内容有如图2所示的对话关系。

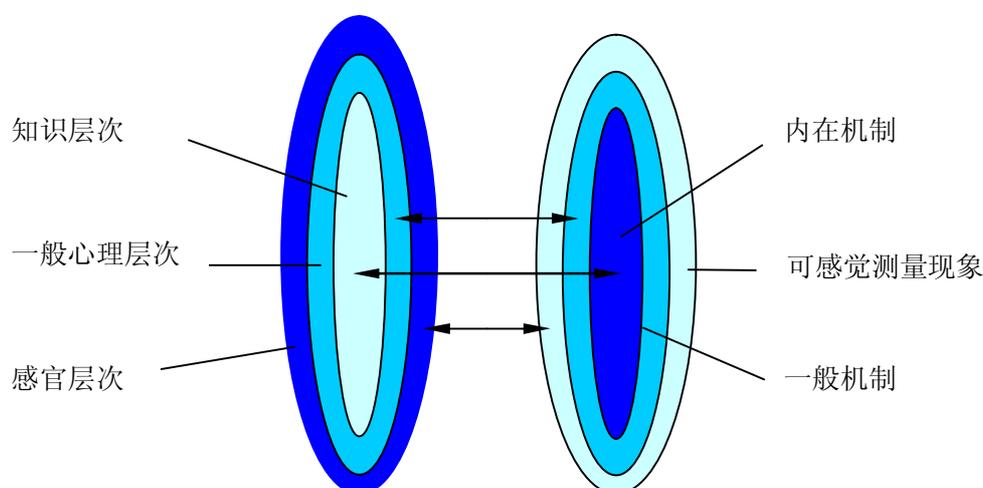


图2 主观认知器与认知对象之间的关系示意图，双向箭头表示认知主体和认知对象各层次之间的对话。

（II）病历数据的生成属性

数据生成器本质上是一种认知器。我们可以假设任何数据都由广义的、被我们概念化的数据生成器生成。而且我们可以进一步假设数据生成器是由广义的观测器和广义的判别（或判断）器组成。图3为从认知对象出发，通过数据生成器到获得数据的认知步骤。对数据生成器的广义观测器组分我们不难理解，对数据生成器必定存在广义的判别器组分的事实我们可以用数据采集的A/D转化作为实例（图4）加以说明。应该说这种假定并未有悖于一般性。

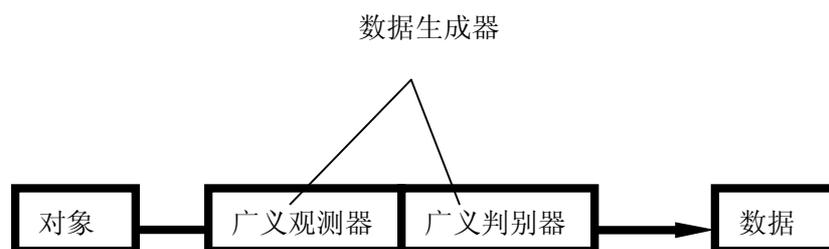


图3 从认知对象通过数据生成器获得数据的认知过程。

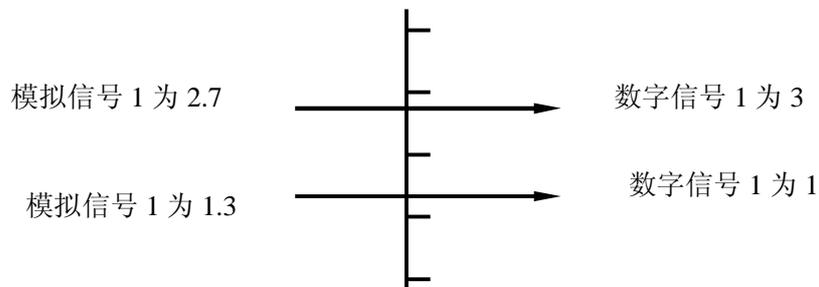


图 4 客观观测中的判断过程实例 (A/D 转换过程中的判断)。

本文把数据生成器分为如下 6 类 (见图 5), 笔者相信不同类型的数据生成器所生成的数据的认知学属性如语义 (覆盖性)、确定性、精确性方面存在差异, 本文称之为数据的生成属性 (generation attribute)。

(1) I 类数据生成器: 如试剂, 仪器和设备等的测量与判断。所生成的数据实例有“体温 37.2℃, 脉搏 76/min, 呼吸 18/min, 血压 15.2/10.1kPa”, “实验室检查血尿素氮 7.14mmol/L, 肌酐 88.4 μmol/L, 血浆总蛋白 55g/L, 血浆白蛋白 28g/L, 血浆球蛋白 27g/L, ALT28U, 血清胆固醇 12.93 mmol/L”。

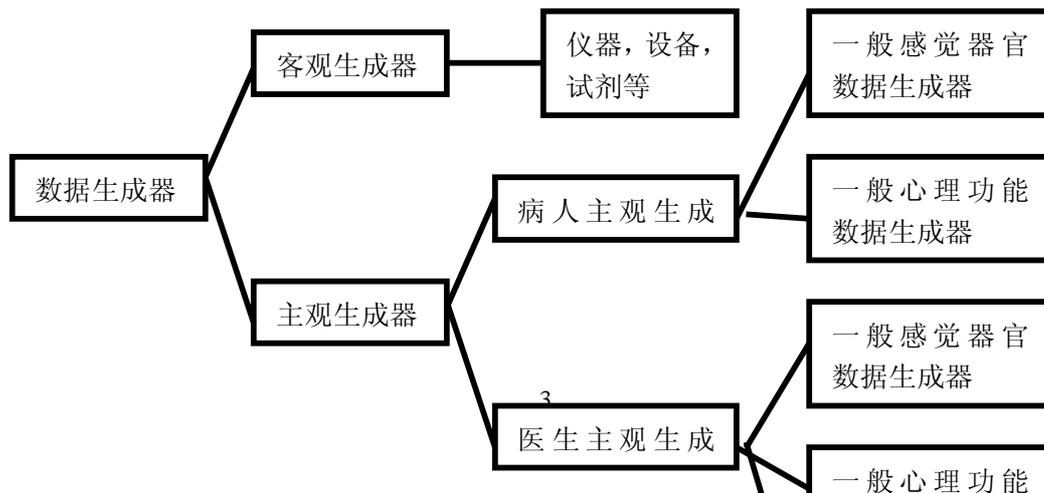
(2) II 类数据生成器: 医生基本感觉层次的测量与判断。所生成的数据实例有“全身皮肤无黄染, 皮疹和紫斑,” “五官未见异常”。

(3) III 类数据生成器: 医生心理层次的测量与判断。所生成的数据实例有“检查欠合作”, “注意力不集中, 对周围持有戒心, 情感反映与思维内容和周围环境不协调”。

(4) IV 类数据生成器: 医生专业知识层次的测量与判断。所生成的数据主要包括如体格检查的数据等。实例有“呼吸运动对称, 语颤两侧正常, 两肺呼吸音清晰, 未闻及干湿罗音,” “心前区无隆起, 心尖搏动在左第 5 肋间, 心界不大,” “肝脾未及, 肝上界在右锁骨中线第 5 肋间”, “无血管杂音”, “无干湿罗音”, “肝、脾、肾、胆囊均未触及”, “莫菲式征阴性”; “甲状腺呈弥漫性对称性 II° 肿大, 质地较柔软、表面光滑, 二极可触及细震颤, 并可闻及血管杂音, ...”

(5) V 类数据生成器: 病人或家属基本感觉层次的测量与判断。所生成的数据实例有“月经史: 14*3-4/28-30, 量中等, 无痛经史, 末次月经时间 1991-11-20”等。

(6) VI 类数据生成器: 病人或家属心理层次的测量与判断。所生成的数据实例有“(患者口述) 3 岁时患‘麻疹’两周痊愈”, “(患者主诉) 否认有其他急性传染病史”。



(III) 病历数据的表达属性（形态属性）

对于一个较为独立的数据或信息单元，一般可以抽象并离解出广义的变量及其广义的值变量。如果更为详尽一些，还应包括这一信息单元存在的“场所”，即信息主体或载体。大多数医学数据由这种三元单元（图 6）及其形形色色的嵌合而成（图 7）。这是数据或信息单元的表达或形态学属性（morphological attribute）中最基本的结构，这种基本结构的分析非常有利于病历数据的各种层次的语义开发。表 1 为病历数据单元的三要素表达实例。

表 1 静态数据单元举例

数据主体	特征变量	变量值
(病人)皮肤	黄染	无
(病人)浅表淋巴结	触及	未
(病人)头颅	畸形	无
(病人)双眼结膜	充血	无
(病人)双眼结膜	水肿	无
(病人)瞳孔	大小	相等
(病人)瞳孔	直径	相等
(病人)瞳孔	对光反射	灵敏
(病人)鼻	通气	良好
(病人)耳	脓性分泌物	无
(病人)口唇	发绀	无
(病人)咽	充血	无
(病人)颈	(张力)	软
(病人)气管	(位置)	稍向右侧偏移
(病人)甲状腺	肿大	不
(病人)颈静脉	怒张	无
(病人)右胸	呼吸动度	明显减弱
(病人)右侧胸廓	(形状)	饱满
(病人)右锁骨中线第 4 肋间下	浊音	有
(病人)右胸	语颤	减弱
(病人)右肺浊音区	呼吸音	消失
(病人)心脏	心尖搏动(位置)	第 5 肋间左锁骨中线外

(病人)心脏	心率	1cm 处 84 次/分
(病人)心脏	心律	齐
(病人)心脏各瓣膜听诊 断区	病理性杂音	未闻及
(病人)腹部	(形态, 质地)	平软
(病人)肝	(大小)	未及
(病人)脾	(大小)	未及
(病人)腹部	压痛	无
(病人)腹部	反跳痛	无
(病人)(腹部)	肠鸣音	存在
(病人)肛门	异常(概括)	无

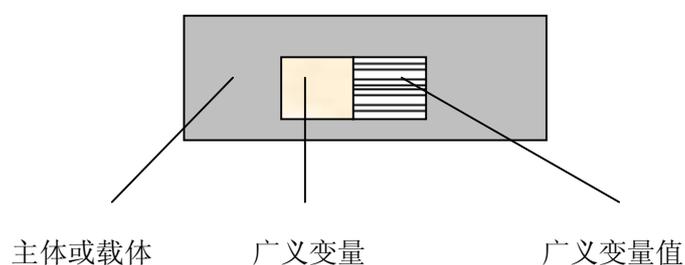


图 6 信息单元由广义主体或载体，广义变量及广义变量值组成。

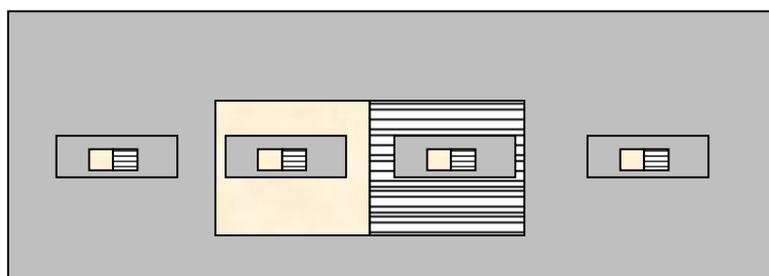


图 7 信息组分的迭代性嵌合示意图：信息单元可以经过复杂的嵌套或分解成为更大的或更小的信息单元。

(IV) 病历数据的认知属性

(i) 病历数据的认知倾向性分析：

病历的陈述性数据的认知目标已有上文给出。一般来说病历数据的认知属性(cognitive attribute)包括如下几方面：

1. 指定性：即语义属性(semantic attribute)，指其各层次的医学(生理，病理等)语义关系。病历数据的语义指定分析可分二步进行：(1) 数据的认知倾向(粗略分析)；(2) 数据的具体物理指定(精细分析)。
2. 确定性(certainty)：为一种数据成立的可靠性(可信度)。(1) 数据本身成立的确定性，由

采集器属性和判断器属性决定；(2) 数据的（临床）语义的确定性，由事物的内在规律决定。数据的确定性决定于认知目标。

3. 粒度 (granularity): 或称精确性, 清晰度。数据的基本粒度也决定于认知目标有关。

数据的具体的物理指定与模式识别领域有关, 已有大量研究, 本文不作讨论。本文首次提出一种所谓粗略分析, 即认知倾向分析。所谓认知倾向是指异常事物的存在性运算(“指定是否存在异常事物”)。它涉及对象空间(由异常事物组成)的发现及其完全性确定和背景空间(干扰空间)完全性清除。如果对象空间的确定是不充分(不完全)的, 就可能遗漏认知目标。没有经历过充分认知操作的背景空间为待测空间或未知空间, 是认知和决策过程的干扰空间, 其中可能潜含对象。因而在没有对背景空间进行充分的认知操作之前, 任何对象空间都是不确定的, 认知目标的实现也是不充分的。传统的病历数据的背景空间查索按各生理系统进行, 即按呼吸、循环、消化、泌尿生殖、神经、精神、血液、五官等系统逐个检查; 也可扩展到对外伤手术史, 中毒和药物过敏史等体外病因空间的检查。

原则讲病历数据的语义“非正常即异常”(也可以把一定的域值归属为中间过渡态), 因此原则上病历数据只有二种认知倾向性: 顺向数据和逆向数据(也许可分出另一种“认知倾向不明”数据)。前者支持“有异常事物存在”, 后者则正相反。数据表示形式直接反映出认知倾向的数据称为显性倾向性数据, 否则称为非显性倾向性数据。

我们用一个简单的例子来说明对象空间、背景空间二者的认知学关系。在 1000 个学生(背景空间)中寻找 A 学生(单元素对象空间), 则发生概率为 $1/1000$; 如把背景空间缩小, 在 100 个学生中寻找 A 学生, 则发生概率为 $1/100$; 在 10 个学生中寻找 A 学生, 则发生概率为 $1/10$ 。而如果我们扩大对象空间, 在 1000 中寻找 A 学生 \vee B 学生(二元素对象空间), 则发生概率为 $2/1000$; 在 1000 中寻找 A 学生 \vee B 学生 \vee C 学生(三元素对象空间), 则发生概率为 $3/1000$ 。这与 Shannon 给出了信息的定量定义 $I = \log_2 P$ 一致。公式中概率 P 越大, 事件信息量越大。如果 P 为 $1/1024$, 则分子 S 相当于对象空间的大小为 1, 分母 M 相当于背景空间的大小(信息容量)为 1024。对象空间(分子)越大, 信息量越大, 反之, 则信息量越小。(当然对于病历数据的认知任务来说, 认知目标是发现所有的异常的对象: 即 $A \text{ 学生} \wedge B \text{ 学生} \wedge C \text{ 学生}$ 。)

病历的顺向数据的认知功能是使对象空间的大小 S_s 足够大, 使之充分分化; 而逆向数据的认知功能是减少背景空间(未知空间, 待测空间)的大小 B_s , 使之最小化。病历数据的认知任务是使 S_s/B_s 趋于最大化(见示意图 8)。

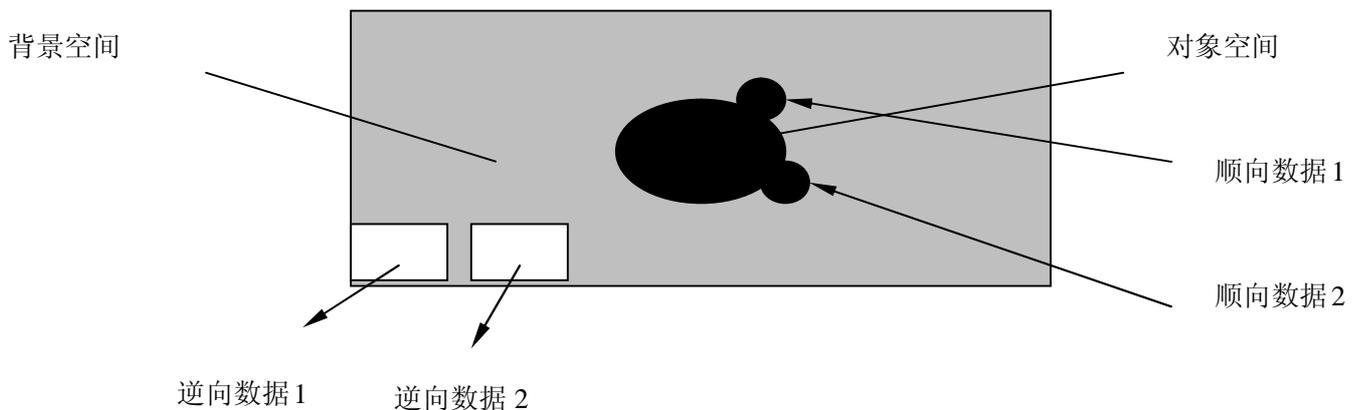


图 8 逆向数据和顺向数据分别通过减少背景(待测)空间的大小和增加对象(异常)空间的大小使病历的背景空间/未知空间最大化。

(一) 显性倾向性数据:

1. 顺向性数据(cognitive forward data)为语义上“支持存在异常事物”的数据。

(1) 肯定性顺向数据以肯定的形式表达一个异常事物或属性,为支持性数据。如“怕热,多汗,乏力,消瘦,心悸,多食,易饥”,“月经紊乱,量少,经期延长”,“甲状腺弥漫性、对称性Ⅱ度肿大,有细震颤及血管杂音,”“ T_4 283.1nmol/L, T_3 4.62nmol/L; 摄 ^{131}I 率显著提高; 心电图显示房颤”。

(2) 否定性顺向数据以否定的形式表达一个异常事物或属性,也为支持性数据。如“右半身活动不灵,语言表达困难,握拳后松开困难 34 年”,“问之不答”。

2. 逆向性数据(cognitive backward data) 为语义上“不支持存在异常事物”的数据。

(1) 肯定性逆向数据:以肯定的形式表达一个正常事物,为不支持性数据。(a) 知识判断型:“眼球运动自如”,“角膜透明”,“呼吸运动两侧对称,节律规则”,“呼吸音和语音传导两侧对称”,“(四肢运动和感觉)良好”,“甲状腺不大,未闻血管杂音”;(b) 如实描述型:“两侧瞳孔对光反射灵敏”,“器官居中”,“(膝腱反射,跟腱反射)存在”。

(2) 否定性逆向数据:以否定的形式表达一个正常事物,也为不支持性数据。(a) 知识判断型:“头颅无畸形”,“无干湿罗音和胸膜摩擦音”。(b) 如实描述型:“皮肤无皮疹,淤斑,淤点,溃疡,疤痕”,“结膜无充血”,“巩膜无黄染”,“口腔黏膜无溃疡”,“角膜无斑翳”,“乳突无压痛”。

(二) 非显性倾向性数据:

此类数据的认知倾向是隐性的或不明的,往往需经有某种语义认知倾向性运算,才能确定其认知倾向。如与语义知识库相关数据比较,才能甄别其认知倾向。如:“双侧肺下界于肩胛下角线第 10 肋间,呼吸移动度 5cm”,“肋弓角约 90° ”,“头发黑”,“舌苔薄白”。

(ii) 逆向性数据的语义指定运算:

由于临床医学认知目标的取向,医学临床知识或病历数据对异常现象及其语义的描述更为详细更为突出,而对异常事物的描述较为简略。这是一种由认知规律决定的数据表达的不对称性。因而顺向性数据往往有明确的临床医学语义,而逆向性知识或病历数据则不然。在这种情况下,一般说来逆向性数据可通过寻找其相应的“反义表述”的顺向性数据,再通过顺向性数据的“临床医学语义”的“反义语义”获得相应的临床医学语义。图 9 为逆向性病历数据“应答切题”获得临床医学语义的过程。

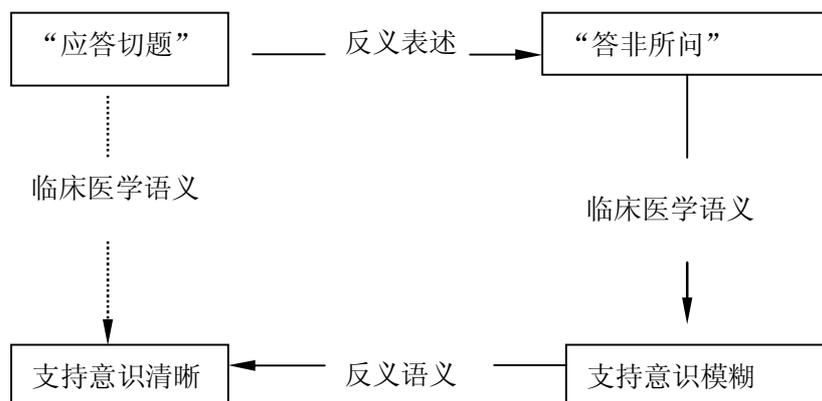


图 9 逆向认知描述(正常描述)的临床医学语义可从正向认知描述(病理描述)的临床医学语义推出。

(V) 再析 BMKI 知识背景理论

随着医学信息学的发展，电子病历将越来越多地涉及支撑性背景知识库。但支撑性知识库的知识不是“拿来就能用”的，这里至少会遇到知识背景空间的问题。BMKI曾着力揭示一条认知科学规律：任何知识的精确度（粒度，即是否足够详尽），确定度，断言能否成立等均决定于其认知背景空间或知识空间。后者又决定于相应的认知目标（意识的、物理的）^[8]。知识背景空间可以是一个公识或公理体系（如欧几里得公理体系），由一些基础命题或规则组成。下面以《医学信息学论坛》^[13]网友提供的一个小故事为例加以说明：

故事是这样的：晚饭后，母亲和女儿一块儿洗碗盘，父亲和儿子在客厅看电视。突然，厨房里传来了一声打破碗的声音，然后一片沉寂。儿子望着他父亲，说道：“一定是妈妈打破的”，“你怎么知道？”“她没有骂人。”——这是故事的物理情景。

故事所确定的认知目标是“判断是谁打破了碗”。要实现这一判断，我们首先把不能停留在物理情景空间，因为物理时空本质上是无穷的。我们不能想象“厨房内可能还有一只猫或一条狗，或碗架不牢固可能坍塌”，或者可能由其他不可测的物理因素实施打破碗的行为。为此我们必须首先实现一个空间转换，把物理情景空间抽象为一个意识空间，后者的特点是“没有提及的事物均不存在”（具有排他性或确定性）。这就是我们的所谓公理体系。由这一认知目标决定的公理体系由如下公识或规则组成：（1）母亲可能打破碗；（2）女儿可能打破碗；（3）母亲会骂女儿；（4）女儿不敢骂母亲。从这个公理系统可以得出二个推理：“（1） \wedge （4）”和“（2） \wedge （3）”，因为母亲“没有骂人”，所以前者是满足物理情景空间的。

具有推理能力的知识背景空间是类似欧几里得空间的公理或公识系统。对学术性知识的一般性理解或联想依赖的知识背景空间为所谓“大或”空间，即很多不同的空间的“或”关系（ $S=S_1 \cup S_2 \cup S_3 \cup \dots$ ），因而不是一个单一的空间。在这一概念下，我们明意识地或下意识地根据大脑工作的需要在不断更换具体的有效知识背景空间。如我们在讲解，阅读或撰写一本医学教科书就是在“大或”空间中进行的。但严密的知识推理只能在同一个公理或公识空间中进行。

现在我们来看由José L.V. Mejino和Cornelius Rosse给出两个推理问题^[14]。“因为前胫骨肌-*contained in*-腿前间隙AND腿前间隙-*part of*-腿，所以前胫骨肌-*part of*-腿。”这个推理的结果断言是正确的。但是下面的推理是无效的：“因为尿-*contained in*-膀胱腔AND膀胱腔-*part of*-膀胱，故尿-*part of*-膀胱。”

Rosse教授给出的解释是：*-part of*-关系必须是两个维度相同的实体之间才能成立，因为尿和膀胱的维度不同，所以“尿-*part of*-膀胱”断言不能成立（Rosse教授给作者的e-mail中阐明了这一意见）。现在我们从“推理要求空间同一性”原理来分析这一问题，也可以得出同样的结论。如果把这两个推理链放在欧氏空间中，两个推理无疑都是正确的。但何以在传统的医学意义（这里 $S=S_p \cup S_a$ ， S_p 和 S_a 分别为传统的生理学公识空间和传统的解剖学公识空间）下，两个看似相似的推理前者成立而后者无效？根本原因在于：前面一个推理是在统一的 S_a 中进行的；而后面一个推理则不然：当我们说“尿-*contained in*-膀胱腔”我们在 S_p 中给出此断言，而当我们说“膀胱腔-*part of*-膀胱”我们是在 S_a 中给出此断言。而在 S_a 中，我们有一条“公识”或“基础命题”，即“**解剖学研究的对象是机体固有的或内在的组成成分**”，因此 S_a 排除了“**临时性储存在机体内的产物**”（但Rosse教授等的FMA的背景知识空间中并没有排除）。不幸的是“尿”正是这种“**临时性储存在机体内的产物**”。既然 S_a 排除了尿这个要素，那么“尿-*part of*-膀胱”断言自然就不能成立。

作为推理基础的公理化空间一般是潜意识存在，它们的建立、显化或科学化可在相应的推理循环过程中逐步实现：

- (1) 在“大或”空间或已有的、尚待完善的公理体系中作出一个推理；
- (2) 检验此推理的有效性：如果“yes”则返回(1)，否则向下；
- (3) 分析不合理的原因，公理体系存在何种定义、公理缺失或冲突？
- (4) 对定义或公理作相应的增减或修改；
- (5) 增减或修改公理体系；
- (6) 返回(1)。

参考文献

1. Bao H.F., Geng J.H. and Su Z.F.: Pansystems Methodology(PM) and a new research on large-scale integration of biomedicine—An introduction of QMSOC and its recent progresses, Acta of Jiansu Industrial College, Journal of Jiangsu Institute of Technology, 1991, 4 (2): 69-75
2. 包含飞：生物医学知识整合论(一)，《医学信息》杂志，16(6)：174-279, 2003
3. 包含飞：生物医学知识整合论(二)，《医学信息》杂志，16(8)：410-415, 2003
4. 包含飞：生物医学知识整合论(三)，《医学信息》杂志，16(11)：602-605, 2003
5. 包含飞：生物医学知识整合论(四)，《医学信息》杂志，16(12)：666-670, 2003
6. 包含飞：生物医学知识整合论(五)，《医学信息》杂志，17(5)：244-250, 2004
7. 包含飞：生物医学知识整合论(六)，《医学信息》杂志，17(8)：452-457, 2004
8. 包含飞：生物医学知识整合论(七)，《医学信息》杂志，17(11)：685-692, 2004
9. Bao H.F.: The Theory of Biomedical Knowledge Integration(I), Chinese Journal of Medical Treatment, 2(13): 1-6, 2003
10. Bao H.F.: The Theory of Biomedical Knowledge Integration(II) ——What is underlying the domain concepts, Chinese Journal of Medical Treatment, 2(21): 1-6, 2003
11. Bao H.F.: The Theory of Biomedical Knowledge Integration(III) ——A Rediscussion on Background Space of Knowledge(BSK), Chinese Journal of Medical Treatment, 3(8): 1-5, 2004
12. 包含飞：生物医学知识整合科学思维的成分异源性(I)，计算机应用与软件，20(增刊)：171-173, 2003
13. 《医学信息学论坛》：<http://www.miforum.net/>
14. José L.V. Mejino, and Cornelius Rosse: Symbolic modeling of structural relationships in the Foundational Model of Anatomy, KR-MED 2004: 48-62