

眼科信息学：数据与知识的整合研究（I）

王晓幸 杨杰 叶良 王勤美 包含飞*

温州医学院附属眼视光医院信息中心 325027, wxx@mail.eye.ac.cn

*上海中医药大学, 201203, bhf2002@online.sh.cn

摘要 本文用大量眼科数据、信息和知识讨论了眼科信息学的知识整合原理和方法, 包括信息单元的载体-变量-变量值三联结构及各种成分的类型分析及其运算, 进而讨论了各类知识特别是群体知识和个体知识的相互转化规律。最后介绍了 BMKI 在理性知识本体、实验性知识本体及罗盘-灯塔式知识本体 (CBO) 的研究进展, 并提出了探索和开发眼科 CBO 本体的设想。

关键词 眼科信息学 生物医学知识整合论 医学信息学 医学知识工程

还原论在创建现代科学大厦并立下丰功伟绩的同时, 把本来应该统一和整体化的知识体系分成了许多很大程度上相互分开或孤立的知识实体。因为严格地说, 这些知识实体只有在其相应的规定环境下才能成立。生物医学知识整合论 (The Theory of Biomedical Knowledge Integration, BMKI)^[1-10]则相反, 它着重各种知识实体的互联和互转化。这不仅是对代表人类的科学事业主流方法学的还原论的一个必要的补充和哲学平衡, 因而有着巨大的理论意义; 对滚滚而来的生物医学知识工程事业, 对这个事业中知识与知识之间相互错综复杂的联系性和影响性, 也有着巨大的实际指导意义。

1. 先解析信息单元

我们的任何认知或科研活动都起始于对某一对象 (信息载体, C) 发生兴趣, 我们于是选中对象某一方面 (特征、行为或与其他事物之间的关系) (信息变量, F) 进行观测 (观察、测量), 并得到一个结果 (变量值, V)。BMKI (生物医学知识整合论) 把这三者称为基本信息单元 (三联体)。

眼科信息学的宗旨是研究眼科学领域中的信息的特征和行为规范, 但它不能离开这些基本信息规律。表 1 为眼科信息三联体单元的实例。

表 1 为眼科信息三联体单元的实例。

| 载体 (C) | 变量 (F) | 变量值 (V) |
|--------|--------|----------|
| 右眼 | 红 | (存在) 3 天 |
| 右眼 | 痛 | (存在) 3 天 |
| 右眼 | 畏光 | (存在) 3 天 |

| | | |
|-----|--------|------------|
| 右眼 | 流泪 | (存在) 3 天 |
| 右眼 | 远视力 | (为) 1.0 |
| 左眼 | 远视力 | (为) 0.8 |
| 右眼 | 近视力 | (为) 0.6 |
| 左眼 | 近视力 | (为) 1.0 |
| 左眼 | 角膜中央厚度 | (为) 500 微米 |
| 右眼 | 肿胀 | 轻度 |
| 球结膜 | 充血 | 中度 |
| 角膜 | 混浊 | (存在) 雾状 |
| 前房 | 深度 | 3.5 毫米 |
| 虹膜 | 纹理 | 不清 |
| 瞳孔 | 直径 | 7.0 毫米 |
| 瞳孔 | 形状 | 竖椭圆形 |
| 瞳孔 | 散大 | 是 |
| 瞳孔 | 对光反射 | 消失 |
| 视盘 | 色泽 | 正常 |
| 视盘 | 边界 | 清楚 |

2. 眼科信息学的变量的基本类型

BMKI 指出任何信息单元的变量其本质都是二个事物或多个事物之间的关系，但我们仍可以按信息变量的惯殊之分、恒变之异、单合之别把变量分为特征（或静态变量）变量(见图 1)。所谓势态变量是一种关系场或潜能，很多情况下，只有在相关的角色因子存在或加入时，关系才会实现。如果我们用 $v = f(y, x)$ 来表示一种势态关系，其中 x, y 为广义的关系角色 1 和 2， f 为关系及其 v 为其描述（变量值），侧势态变量实例见表 2。

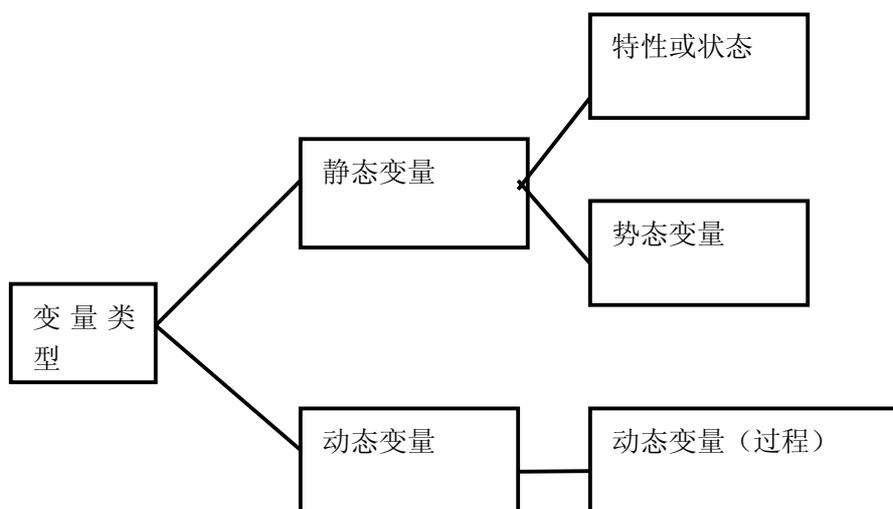


图 1 信息单元的变量类型。

表2 势态关系变量由关系角色 (x,y), 关系 (f), 变量值 (v) 组成。

| 关系角色 1 (x) | 关系角色 1 (y) | 关系 (f) | 变量值 (v) |
|-------------|------------|----------|-------------|
| 上穹隆结膜 | 异物 | 存留 (存在) | 未见 |
| 角膜上皮 | 角膜下方 | 脱落(from) | 大片 (脱落) |
| 直接眼底镜 | 眼底 | 观察 | 不能 |
| 遮盖左眼 | 外周侧假像 | 导致 | 消失 |
| 右眼瞳孔直径 | 左眼瞳孔直径 | 相等 | 成立 |
| 晶状体后囊膜下混浊程度 | 视力受影响的程度 | 比例 | 不成立 |
| 外耳道 | 冰水(50ml) | 注入 | 双眼向注水侧作同向偏斜 |
| 外耳道 | 冰水(50ml) | 注入 | 向对侧的快速眼球震颤 |
| 高眼压 | 降压药物 | 由...治疗 | 缓解 |

动态变量可视为静态变量的时序复合, 在眼科学中可举例如下:

“以往在阅读报刊或看电视稍久后, 常感右眼酸胀、视物不清、额部轻痛, 休息后症状消失”;

“眼睛朝受损外眼肌作用方向注视, 真像与假像间的距离增大”;

“视网膜脱离的先兆可有一阵“火花”或闪光并可出现飘浮物, 只有当视网膜真正和其下面的组织 (视网膜色素上皮) 分离后才会视野中产生帷幕状的视觉缺损区”;

“老年性白内障的主要症状是进行性、无痛性视力减退”;

眼科常见病青光眼的病程常常是一种过程或动态变量:

患者 8 个月前某晚 20:00 左右无情绪激动后出现左眼发红, 眼球呈剧烈胀痛伴同侧头痛, 伴恶心呕吐, 视力急剧下降。在当地医院诊为“血管神经性头痛”对症治疗, 三天后右眼发病, 症状同左侧检查, 在当地医院眼科会诊后发现“左眼压达 51 mmHg”, 确诊为“急性闭角型青光眼”, 予降眼压药物 (具体不详) 治疗 3 天后缓解。7 月余前来我院进一步诊治, 门诊诊为“双眼原发性急性闭角型青光眼 (左眼缓解期, 右眼临床前期)”, 予双眼激光虹膜周切术, 术后情况良好。3 天前左眼再次出现胀痛, 检查发现“左眼压 46.9 mmHg”, 予“乙酰唑胺片、真瑞、美开朗眼水”治疗后缓解。现来我院要求手术治疗。

3. 变量值类型

变量类型可以是数值型(如测眼压右眼 6.6kPa, 左眼 2.7kPa), 符号文字型, 图像型, 音频视频型,但在信息学操作方面一般可以分为可运算型, 可推导性及可理解型。很多变量的变量值仅仅表示“存在”。

4. 信息单元运算或操作

设我们有标准知识语义库 $S = \{(x,y,f)\}$, x,y 为两个变量或两个载体, f 为语义关系(一

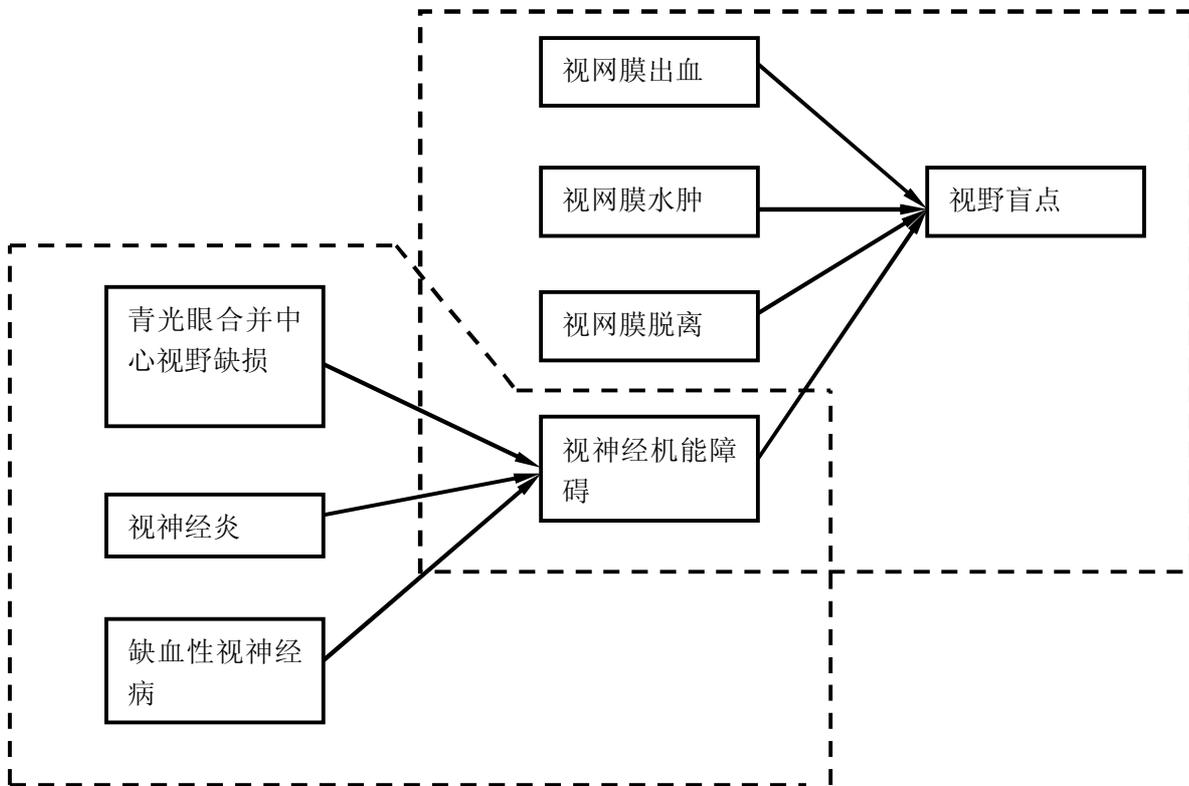


图3 另一例生理-病理型概念性数字虚拟标准眼（部分）。图中箭头表示--导致—关系。由虚线包围的两个子网按认知科学属性均属候选式子网。

(2) 群体信息载体知识和个体信息载体知识的转化:

群体信息载体知识和个体信息载体知识的转化是生物医学认知科学的极为重要内容，群体知识向个体知识转化涉及诸如电子病历或临床及分子遗传学数据库知识工程如聚类分析、模式提取或发现、一致性平均、统计方法等；反方向转化涉及诊断，分类分析，模式识别等。

用统计学方法从个体知识获得群体知识的实例见图4。

实例性知识: 患者8个月前某晚20:00左右无情绪激动后出现左眼发红, 眼球呈剧烈胀痛伴同侧头痛, 伴恶心呕吐, 视力急剧下降。在当地医院诊为“血管神经性头痛”对症治疗, 三天后右眼发病, 症状同左侧检查, 在当地医院眼科会诊后发现“左眼压达 51 mmHg”, 确诊为“左眼急性闭角型青光眼”, 予降眼压药物治疗 3 天后缓解。现来我院要求手术治疗。

统计学方法

标准知识:急性闭角型青光眼急性发作最常见的症状为眼球疼痛, 多为胀痛, 伴混合充血和急剧视力下降, 以及同侧剧烈头痛, 可有恶心、呕吐。诱因包括大量饮水, 长时间处于暗环境下等等。检查可发现前房浅, 房角部分或大部关闭, 多数眼压高达 50 mmHg 以上, 降压治疗可缓解症状, 但病程具有反复性, 可反复急性大发作, 造成永久性的房角粘连和视神经功能损害。

图 4 举例说明关于个体信息载体的知识转化为涉及群体信息载体的知识过程。

(3) 关于群体信息载体的知识——医学本体学

自明斯基提出知识表示的框架理论以来, 框架成为人工智能领域中知识表示的主流方法。目前医学信息学的知识工程技术基本上追随这一潮流, 引导了开发医学知识本体(ontology)的时代的到来。领域本体被称为领域的概念模型, 实际上是一个严格结构化(主要以框架形式表示)的概念属性和关系结构或体系, 本体强调知识单元的重复应用性和共享性。所以它们必须是具有标准意义的群体知识。近来发展趋势是语义 web(semantic web)。具有 Internet 互指性, 所以有全世界统一互补性开发的特点。有些甚至提出“人类的大脑”(注意不是“人的大脑”), 提示将来的工作有人们分工协作, 互补累加的特点。这一令人难以想象的大举措值得相关学者密切关注。下文简单介绍一下语义 web。

语义Web(Semantic Web)作为当前万维网的扩展, 于1999年由Tim Berners-Lee等人提出, 其目的是通过结构化和形式化的手段表示Web上的资源, 不仅能显示信息资源, 而且应用程序能对这些资源进行自动分析和推理。使机器更“理解”网络资源。Tim Berners-Lee于是又在2000年提出了语义Web的体系结构(如图5所示)。其中, 最底层是统一资源标志符(Uniform Resource Identifier, URI)和统一字符编码(Unicode), 为Web上资源标识和字符表示的标准。上一层为可扩展性标识语言XML和定义XML语法结构的XML Schema, 以及允许在文档中合成不同词汇的XML命名空间(Name Space, NS)。这些只是语义Web的语法基础。语义Web体系结构的语义层则从第三层资源描述框架(Resource Description Framework, RDF)开始。后者的形式为(资源 属性 值)三元组, 但因为没有描述属性或说明属性与其他资源之间的关系, 因此需要RDF的词汇描述语言RDF Schema(RDFS)进一步限定和细化。RDFS是在RDF的基础上引进了类的概念、类之间的包含关系、属性之间的包含关系, 以及属性的定义

域和值域。为了进一步满足描述各业务领域的专业知识概念模型的需要，人们又提出本体语言这一表达工具，从SHOE, DAML-ONT, OIL, DAML+OIL, 到最新的OWL(Web Ontology Language)反映了本体语言的发展过程。

眼科信息学的知识本体研究实例见参考文献^[12-13]，所用的本体语言为DAML+OIL。

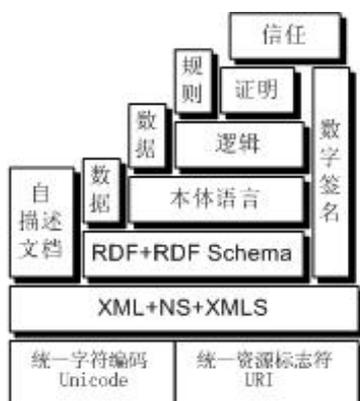


图 5 语义 web 的体系结构。

(4) 非理性或映射性医学本体

基于对当前知识本体的标准性知识的特点。BMKI 引进另一种“标准的、群体的、平均性”知识“本体”，即非理性（实验性）映射性本体（EO），以补足当前生物医学知识工程的标准知识领域的不完整性。EO 被论文^[7]称为“生物算子”的函数或映射就是这类事物。例如图 6。

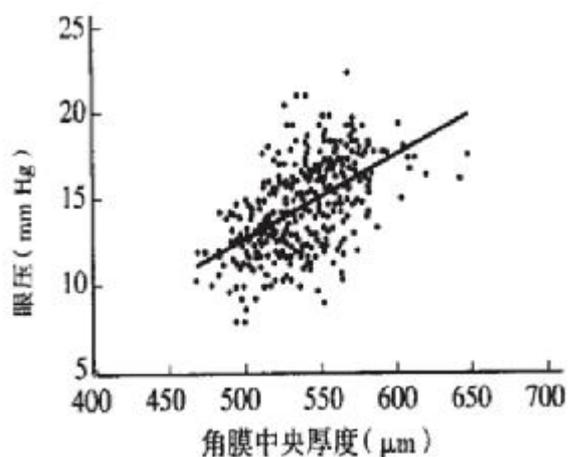


图 6 近视眼患者角膜中央厚度与眼压的关系。以角膜中央厚度为自变量，眼压为应变量，得出直线回归方程 $Y=-10.34 \pm 0.05X$ 。

6. CB 型眼科本体——最接近于真实的本体

如果把CO看作是生物医学知识整合或其他知识工程“大楼的钢筋水泥架构”，属于“粗线条”工程，而又因为临床医学面对的是个体化对象，所以如何使指导性应用程序和知识库的行为尽可能地“体贴入微”，对于标准知识的应用领域来说，我们面前还存在广大的未知领域。上文BMKI提出的非理性EO就是这种待开发领域。

鉴于生物机体的高度复杂性，单独通过纯理性推导、运算等手段达到生物医学知识整合无疑是不可能的，于是问题就又回到BMKI提出的罗盘-灯塔式本体(Compass-beacon Ontology, CBO)^[8]，它是RO与EO的结合。

因此，CBO是医学信息学本体也是眼科信息学本体的更为完备、更为细腻的，也更为有力的标准知识实体。而如果与大量的个体信息（电子病历、大量的数据库）结合起来，我们将面临眼科信息学知识工程的最为完美的挑战。

参考文献

- 1.包含飞：生物医学知识整合论(I)，《医学信息》杂志，16(6)：174-279，2003
- 2.包含飞：生物医学知识整合论(II)，《医学信息》杂志，16(8)：410-415，2003
- 3.包含飞：生物医学知识整合论(III)，《医学信息》杂志，16(11)：602-605,2003
- 4.包含飞：生物医学知识整合论(IV)，《医学信息》杂志，16(12)：666-670,2003
- 5.包含飞：生物医学知识整合论(V)，《医学信息》杂志，17(5)：244-250,2004
- 6.包含飞：生物医学知识整合论(VI)，《医学信息》杂志，17(8)：452-457,2004
- 7.包含飞：生物医学知识整合论(VII)，《医学信息》杂志，17(11)：685-692,2004
- 8.包含飞：生物医学知识整合论(VIII)，《医学信息》杂志，18(1)：11-15，2005
- 9.包含飞：生物医学知识整合论(IX)，《医学信息》杂志，18(3)：172-176，2005
- 10.包含飞：生物医学知识整合论(X)，《医学信息》杂志，(待发表)
- 11.包含飞：MBKI的概念，原理和方法学(I)，(待发表)
- 12.梅婧 刘升平 林作：Chpt-63-语义Web语言的逻辑分析，<http://www.is.pku.edu.cn/~lz/publications/semanticweblang.pdf>
- 13.M. Taboada¹, M. Argüello¹, J. Des² and D. Martínez¹, 包含飞,王晓幸翻译：DAML+OIL 应用实例：眼科学领域的一个本体，<http://www.miforum.net/bhf/ykbt.doc>