

叶 鹰

抽象图书馆学的研究方法

ABSTRACT The key of the methodology of abstract library science is to introduce modern scientific, especially mathematical, methods into the study of library science. The author proposes some ideas of establishing analytical models of abstract library science and explains methodology of abstract library science by solving some specific forms of abstract library equations. 3 refs. 35 formulas.

KEY WORDS Abstract library science. Methodology in library science. Library analysis.

CLASS NUMBER G250

笔者在文献[1]中已经提出了抽象图书馆的定义和抽象图书馆学的基本思想, 本文将继续对抽象图书馆学的研究方法作深入研讨。

1 核心思想及分析模型

抽象图书馆学研究方法的核心思想是将现代科学研究方法尤其是数理方法引入图书馆学研究, 使图书馆学研究趋于精密化、规范化, 进而在科学界奠定其牢固的学术基础。

现代科学研究的基本方法首先是分析, 因而抽象图书馆学研究也应从建立抽象图书馆分析模型入手。

从抽象图书馆基本方程

$$L(I, i) = f[N(K, T)] \quad (1)$$

或

$$N(K, T) = F[L(I, i)] \quad (2)$$

中, 应该找出知识水平 K 、技术水平 T 与抽象图书馆输入信息 i 和输出信息 I 之间的具体定量关系, 才能进行量化分析。

于是, 寻求信息与知识之间的关系成为关键问题。著名英国情报学家 Brookes 曾考虑过该问题^[2,3], 但他提出的新知识 (NK) 由已有知识 (K)

加上新增信息 (dI) 构成, 并不能解决新增知识 (dK) 与新增信息 (dI) 的关系问题, 而从逻辑上我们也只能说新知识 (NK) 应由已有知识 (K) 加上新增知识 (dK) 构成 ($NK = K + dK$), 故在已有理论中不能找到现成答案。现从一般意义上进行分析: 一方面, 信息与知识之间是有区别的; 另一方面, 信息与知识又是相互联系并可以相互转化的。由此作为思考切入点, 可望发现量化线索。

信息与知识的区别是显然的, 可以举出通常是知识而不是信息的例子, 如: $3 \times 3 = 9$; 地球绕日运行, 也可以举出通常是信息而不是知识的例子, 如: 下雨了; 涨价了。同时在表达形式上, 知识常用判断句、规则(若...则..., if...then) 语句和语义网络、公式等表达; 而信息常用存在句、陈述句、图表、数据等表达; 在实质内容上, 知识一般具有内在结构, 而信息则主要体现在其具有的功能上。这种区别类似人工智能和专家系统设计中知识库与数据库(信息库)的区别。

但信息与知识又是相互联系的, 要理解吸收信息必需一定的知识基础, 如果不知道“价格上涨”的含义便无从理解“涨价了”这一信息。同时, 知识一般不成为信息, 尤其是已成为人类公有知识的常识。但在特定情况下, 知识也可能转化为信息, 如约定以“九九八十一”来暗示成功, 则“九

九八十一”就传达了“成功”这一信息。从信息积累中找出规律性,就可以构成知识。

可以考虑的思路是:转化为知识的信息是有效信息,只有有效信息才能转化成知识,无效和冗余信息则无此功能。这样,设有效信息为 J , 则 J 可以引入下式度量:

$$J = I^v \quad (3)$$

其中 $v \in [0, 1]$ 是信息的价值系数,当 $v = 1$ 时, $J = I$, 所有信息都是有价值的;而当 $v = 0$ 时, $J = 1$, 可以认为只告诉了有信息存在这一消息;

$v \in (0, 1)$ 时,一些特殊值恰好与 Rescher 模型和 Rossean 定律(重要文献是文献总数的平方根)对应:

	v	J
常规信息:	$v = 1$	$J = I$
有用信息:	$v = 3/4$	$J = I^{3/4}$
重要信息:	$v = 1/2$	$J = I^{1/2}$
非常重要信息:	$v = 1/4$	$J = I^{1/4}$
最重要信息:	$v = 0$	$J = \ln I$ (定义)

一般情况下, v 相当于价值百分比系数, v 越大表明有效信息越多。传递着信息是 I , 而真正有用的只是 J 。($I - J$) 即为信息冗余。

从概念上讲,有价值的信息才会使知识增加,故引入“单位信息增量产生的单位知识增量应与有价值信息量(有效信息量)成正比”这一关系是合理的,也兼顾了数学上的简单性和信息与知识含义的复合性,即:

$$\frac{dK}{dI} = J \quad (4)$$

用 δ 作为信息的知识转化系数,则

$$\frac{dK}{dI} = \delta I = \delta I^v \quad (5)$$

此时 K 的计量单位与 I 的计量单位一致。

由(5)对 I 求二阶导数得:

$$\frac{d^2K}{dI^2} = \delta v I^{v-1} \quad (6)$$

因(6) - $\frac{v}{I}$ (5) = 0, 故有:

$$\frac{d^2K}{dI^2} - \frac{v}{I} \frac{dK}{dI} = 0 \quad (7)$$

这是在未考虑图书馆内存在信息源的情况下获

得的知识与信息之间的动力关系方程。当考虑图书馆本身也是一个信息源时,仿照系统中存在“动力”或“源”的数学模型,应在(7)式右端增加一个源函数 $f(K)$, 成为:

$$\frac{d^2K}{dI^2} - \frac{v}{I} \frac{dK}{dI} = f(K) \quad (8)$$

(8)应为图书馆知识与图书馆输出信息之间的一般方程,可称为抽象图书馆第一分析方程。

同时,考虑图书馆技术水平与信息的关系时,可以认为信息处理能力实质上代表着技术水平,这样,图书馆技术水平应与单位时间内图书馆对输入信息的处理量成正比,即:

$$T = \frac{di}{dt} \quad (9)$$

引入图书馆的技术转化(或信息消化)系数 ϵ , 得:

$$T = \epsilon \frac{di}{dt} \quad (10)$$

其中 T 无量纲, ϵ 计量单位为 s/byte, i 计量单位为 byte, t 计量单位为秒(s)。将(10)对 t 求二阶导数,得:

$$\frac{d^2i}{dt^2} + \left(\frac{1}{\epsilon} \frac{d\epsilon}{dt} \right) \frac{di}{dt} = \frac{1}{\epsilon} \frac{dT}{dt} \quad (11)$$

(11)应为输入信息与图书馆技术水平之间的一般方程,可称为抽象图书馆第二分析方程。

方程(8)和方程(11)在数学上均为变系数二阶线性常微分方程。

2 抽象图书馆方程的分析解及其意义

这里并不试图求出方程(8)和(11)的一般分析解,而是仅对特定条件下的分析解作一讨论。

(8)式中,当 $f(K) = \frac{m^2}{I^2} K$ 形式时,令 $x = \ln I$, 则因有

$$\frac{dK}{dI} = \frac{dK}{dx} \frac{dx}{dI} = \frac{1}{I} \frac{dK}{dx} \quad (12)$$

$$\frac{d^2K}{dI^2} = \frac{d}{dI} \left(\frac{1}{I} \frac{dK}{dx} \right) = \frac{1}{I^2} \frac{d^2K}{dx^2} - \frac{1}{I^2} \frac{dK}{dx} \quad (13)$$

故可将(8)变换为常系数二阶线性微分方程:

$$\frac{d^2K}{dx^2} - (1 + v) \frac{dK}{dx} - m^2 K = 0 \quad (14)$$

(14)的特征方程为 $r^2 - (1+V)r - m^2 = 0$, 其解为二实根

$$r_1 = \frac{(1+V) + \sqrt{(1+V)^2 + 4m^2}}{2},$$

$$r_2 = \frac{(1+V) - \sqrt{(1+V)^2 + 4m^2}}{2},$$

于是(14)的通解为:

$$K = C_1 \exp[r_1 x] + C_2 \exp[r_2 x] \quad (15)$$

其中 C_1 和 C_2 为两积分常数. 将 $x = \ln I$ 代回(15), 得:

$$K = C_1 I^{r_1} + C_2 I^{r_2} \quad (16)$$

这就是图书馆知识 K 与图书馆输出信息 I 之间的一个分析解, 它表明图书馆知识是图书馆输出信息的复合价值指数和形式.

同样, 对(11), 当 $\frac{1}{\epsilon} \frac{d\epsilon}{dt} = p$ 和 $\frac{1}{\epsilon} \frac{dT}{dt} = qt$ 形式且 p, q 为常数时, (11)化为常系数二阶线性微分方程:

$$\frac{d^2 i}{dt^2} + p \frac{di}{dt} - qt = 0 \quad (17)$$

其通解为:

$$i = C_1 \exp\left[\frac{-p + \sqrt{p^2 + 4qt}}{2} t\right] + C_2 \exp\left[\frac{-p - \sqrt{p^2 + 4qt}}{2} t\right] \quad (18)$$

这是抽象图书馆输入信息 i 与技术水平 T 之间的一个分析解, 它表明抽象图书馆技术水平以指数形式影响信息处理.

3 传统图书馆的抽象分析

对一般抽象图书馆来说, 直接用上述方程分析较有普遍意义, 而对传统图书馆来说, 由于其特殊性, 某些分析还可简化和具体化. 首先, 传统图书馆的技术水平 T 可以还原为反映其内部构成和性能的人力、财力、物力指标, 即可用人、财、物资源数量的加权计量将 T 表示为:

$$T = aA + bB + cC \quad (19)$$

其中 $a + b + c = 1$ 是人、财、物权重, 而

$$A = \sum_i a_i A_i \quad (20)$$

是图书馆内各类人员的技术评分加权值, a_i

$= 1$ 为其权重;

$$B = \sum_i b_i B_i \quad (21)$$

是图书馆内各类财产的技术评分加权值, $b_i = 1$ 为其权重;

$$C = \sum_i c_i C_i \quad (22)$$

是图书馆内各类设备的技术评分加权值, $c_i = 1$ 为其权重.

其中 A, B, C 的具体计算方法可以设计如下:

对一个拥有正高职人数为 A_1 、副高职人数为 A_2 、中职人数为 A_3 、初职人数为 A_4 、博士人数为 A_5 、硕士人数为 A_6 、学士人数为 A_7 、专科人数为 A_8 的图书馆来说, A 可以设计为:

$$A = 0.20A_1 + 0.15A_2 + 0.10A_3 + 0.05A_4 + 0.20A_5 + 0.15A_6 + 0.10A_7 + 0.05A_8 \quad (23)$$

对购电子出版物经费为 B_1 、购传统书刊经费为 B_2 、运行经费为 B_3 的图书馆, B 可以设计为:

$$B = 0.5B_1 + 0.3B_2 + 0.2B_3 \quad (24)$$

对价值5万以上设备数量为 C_1 、价值1万~5万设备数量为 C_2 、价值5千~1万设备数量为 C_3 、价值1千~5千设备数量为 C_4 的图书馆, C 可以设计为:

$$C = 0.4C_1 + 0.3C_2 + 0.2C_3 + 0.1C_4 \quad (25)$$

对较重视人的作用的图书馆, 可以取

$$a = 0.5, b = 0.25, c = 0.25;$$

对较重视设备的图书馆, 可以取

$$a = 0.3, b = 0.25, c = 0.5;$$

对各方面均衡考虑的图书馆, 可以取

$$a = 0.35, b = 0.3, c = 0.35.$$

这样, 用一套统一的指标测算同一时期某批图书馆时, 就可以比较出各馆技术水平 ($T = aA + bB + cC$) 的高低和信息处理能力 ($di/dt = T/\epsilon$) 的强弱. 通过改换不同的 a, b, c , 可以找出影响同一批传统图书馆技术水平高低的主导因素是人员、资金还是设备, 进而可以找到解决问题的途径.

如果采用计算机系统地进行计算处理, 还可

以引入仿真方法。

此外,根据图书馆函数的定义,有:

$$L = \frac{I}{i} = \frac{\frac{1}{\epsilon} \left(\frac{dK}{dI} \right)^v}{\frac{1}{\epsilon} T dt} \quad (26)$$

故

$$\frac{dK}{dI} = \left(\frac{\delta}{\epsilon} \frac{I}{i} T dt \right)^v \quad (27)$$

即

$$K = \left(\frac{\delta}{\epsilon} \frac{I}{i} T dt \right)^v dI \quad (28)$$

这表明 T 同时也间接地决定了图书馆的知识水平。

(28)与(5)的积分形式

$$K = \delta I^v dI \quad (29)$$

是兼容的,(29)表明知识是有效信息的积分。

4 抽象图书馆的优化

从图书馆函数和图书馆方程的数学形式看,抽象图书馆的优化可分解为知识优化和技术优化两部分。

对于知识优化,由于优化点应出现在 $\frac{d^2K}{dI^2} = 0$ 处(极值原理),故由(8)有:

$$\frac{dK}{dI} = - \frac{I}{V} f(K) \quad (30)$$

对 $f(K) = \frac{m^2}{I^2} K$ 情形,上式经分离变量后成为:

$$\frac{dK}{K} = - \frac{m^2}{V} \frac{dI}{I} \quad (31)$$

两边积分得:

$$\ln K = - \frac{m^2}{V} \ln I + \ln C \quad (32)$$

C 为积分常数,即:

$$K = I^{-\frac{m^2}{V}} C \quad (33)$$

于是获得一系列优化点。

对于技术优化,同样当 $\frac{d^2i}{dt^2} = 0$ 时,由(17)有:

$$\frac{di}{dt} = \frac{q}{p} t \quad (34)$$

解得:

$$i = \frac{q}{p} t dt + C = \frac{q}{2p} t^2 + C \quad (35)$$

C 为积分常数,这表明 $i = \frac{q}{2p} t^2 + C$ 时存在一系列优化点。

用同样方法还应能更进一步分析抽象图书馆的结构优化、功能优化、资源优化等更深的问题,但必须从图书馆函数和图书馆方程着手寻求。

5 小结

以上公式大体展示了抽象图书馆学研究方法的概貌,通过这些研究方法,使抽象图书馆学的思想与实际应用之间贯通,进而使抽象理论得以产生实用价值。

显然,从不同的抽象图书馆的分析模型出发必然会得到不同的分析结论。本文所述仅是有代表性的其中一种情形,按照同样思想与方法,读者完全可以根据实际需要建立自己独特的分析模型,进而获得有新意的分析结论,这就是抽象图书馆学方法为图书馆学理论建设指出的丰富多彩的前景。

希望抽象图书馆学的思想和方法能为当今图书馆学理论注入生机和活力,使图书馆学研究走上精密科学的发展道路。

参考文献

- 1 叶鹰 图书馆学基础理论的抽象建构 中国图书馆学报, 1998, 24(3): 86~ 88
- 2 Brookes, B. C. The Fundamental Equation of Information Science in Research on the Theoretical Basis of Information. Moscow: International Federation for Documentation, 1975
- 3 Brookes, B. C. The Foundation of Information Science Journal of Information Science, 1980, 2 (3~ 4): 125~ 133; 1980, 2(6): 269~ 275; 1981, 3(1): 3~ 12

叶鹰 哲学博士。现任浙江大学图书馆副馆长、浙江大学文献信息研究所副所长,副研究馆员。
通讯地址: 杭州市。邮编 310027。

(来稿时间: 1998-08-11, 编发者: 徐苇)