第 18 卷 第 2 期 湖南科技大学学报(社会科学版) Vol. 18 No. 2
2015 年 3 月 Journal of Hunan University of Science & Technology(Social Science Edition) Mar. 2015

doi:10.13582/j.cnki.1672-7835.2015.02.005

2^n 值命题演算的语义及其在大数据中的应用①

洪龙[1,2]

(1. 南京邮电大学 计算机学院, 江苏 南京 210003;2. 软件开发环境国家重点实验室, 北京 100191)

摘 要:本文的目标是建立 2^n 值命题演算(2^n - Valued Propositional Calculus, 2^nP)的语义,为大数据科学奠定逻辑基础。描述了 2^n 值逻辑的真值形式、采用位结构刻画了联结词的功能,并建立了赋值映射;根据多值逻辑的特点,以数据冗余、key - value 模型为例,直觉地讨论了 2^n 值逻辑应用于大数据研究的有效性;初步分析了 2^nP 语义与经典命题逻辑语义之间的关系,并展望了 2^nP 在计算机科学、人工智能、信息技术等学科的应用前景。

关键词:命题演算;语义;大数据;降冗;key - value 模型

中图分类号:B81 **文献标志码**:A **文章编号**:1672 - 7835(2015)02 - 0023 - 05

The Semantic of 2^n - Valued Propositional Calculus and Its Applications in Big Data

HONG Long[1, 2]

(1. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;
2. State Key Laboratory of Software Development Environment, Beijing 100191, China)

Abstract: This paper aims to establish the semantic of 2^n valued propositional calculus (2^nP) in order to set the logical foundation of big data science. This paper specifies the truth value form of 2^n valued logic and uses bit structure to characterize the functions of connectives, and set valuation mapping of 2^nP. In light of the features of 2^nP and by taking key - value data model and data redundancy as examples, this paper intuitively discusses the effects of 2^nP applied to the researches on big data. Furthermore, this paper initially analyzes the relationship between semantic of classic propositional logic and the one of 2^nP, and shows the 2^nP's application perspectives on computer science, information techniques, and artificial intelligence.

Key words: proposition calculus; semantic; big data; redundancy cancelling; key - value model

引论

大数据现已成为当今社会的热词。由于它具有潜在的巨大价值,是未来的“新石油”,许多国家的基金机构对其研究给以支持,从而使得大数据演化为大科学[1][2]。

大数据的“大”体现在三个方面:系统必须采样、处理和传播的信息量;要处理的信息类型的数量和复杂性;信息流的速率。据著名咨询公司 IDC 的统计,2011 年全球被创建和复制的数据总量为 1.8ZB(1.8×10^{21} 字节),其中 75% 来自于个人(主要是图片、视频和音乐),远远超过人类有史以来所有印刷材料的数据总量[3]。尽管大数据中蕴含着不可估量的价值,但由于数据规模巨大,从何处得到金矿则

---

① **收稿日期**:2014 - 08 - 03

基金项目:国家自然科学基金项目(61170322);软件开发环境国家重点实验室开放课题(SKLSDE - 2011KF - 04)

作者简介:洪龙 (1952 -),男,江苏南京人,教授,博士,主要从事非经典逻辑及其应用、计算机科学和人工智能研究。

是我们面临的难题。从另一方面看,若存在海量的、无序的,特别是内容相同的数据,则在我们挖掘知识时,它们会成为障碍,甚至是灾难[4]。

面对这些被称作数据山[5]、数据洪水[6]的数据和大数据潜在价值的巨大诱惑,人们期盼数据科学[7]的到来。在科学研究中,定性分析是不可缺少的环节,而逻辑真值的本质作用就是进行定性。因此构建一种适用于大数据的逻辑,对于建立大数据科学具有奠基作用。

在计算机科学和信息技术中,数据是以位或字节为单位进行组织、存储和处理的,数据的量是以 2^i 表示的。可以直觉感到:本文将要研究的 2^n 值逻辑与具有 2^m 字节的大数据之间有某种关联, 并可能在大数据的研究中发挥作用,这里 $n < m$。

2^n 值逻辑属于多值逻辑。1920 年,Lukasiewicz 提出三值逻辑并称其是第一个非 Aristotle 的逻辑[8],1921 年,Post 在研究媒体演示的一般理论时独立地创建了多值逻辑[9]。两位伟大的科学家突破了经典逻辑的二值束缚,为自然地描述“非此非彼”奠定了基础。

本文的目标是建立既保留经典逻辑语义模型,又体现自身特征的 2^n 值命题演算(2^n - Value Propositional Calculus, 2^nP)的语义框架。需要说明,对本文的启示来自于两个方面。

(1) Post 在描述 Principia 数学系统的连接词时首先引入了否定词和析取词,然后以它们作为基础定义其它连接词[9]。这是一种由简至繁的、自然的逻辑系统的构建过程,具有方法论意义。

(2) Belnap 在论述蕴涵与相关时认为:A 蕴涵 B 的前提必须是两者相关联,而 A 与 B 相关联的必要条件是有共同的逻辑变量,他还给出了符合该条件的真值表[10]。

余下的部分这样安排:第 2 节论述 2^nP 的真值和联结词,这是本文的主要内容之一;第 3 节介绍赋值,探讨 2^nP 应用于大数据的有效性在第 4 节展开,并在第 5 节总结全文。

1 真值与联结词

记 2^n 值逻辑的真值形式为 $x_{n-1}x_{n-2}\cdots x_0$. 其中 $n \in N$, $x_i \in \{0, 1\}$, $0 \leqslant i \leqslant n-1$. 并记 2^n 值逻辑真值的集合为 S_{2n}. 显然

$$S_{2n} = \{2^n-1, 2^n-2, \cdots, 1, 0\} = \{11\cdots1, \underbrace{1\cdots10}_{n\ \text{bits}}, \cdots, 0\cdots01, 0\cdots00\}.$$

定义 1 设 $x, y \in S_{2n}$, $x = x_{n-1}x_{n-2}\cdots x_0$, $y = y_{n-1}y_{n-2}\cdots y_0$.

(1) 符号¬ 称作否定联结词, 若 $\neg x = {}^{\neg}x_{n-1}{}^{\neg}x_{n-2}\cdots{}^{\neg}x_0$,其中 ${}^{\neg}x_i = \begin{cases}1 & x_i = 0\\ 0 & x_i = 1\end{cases}$.

(2) 符号∨称作析取联结词. 若 $x \vee y = {}^{\vee}x_{n-1}y_{n-1}{}^{\vee}x_{n-2}y_{n-2}\cdots{}^{\vee}x_0y_0$,其中 ${}^{\vee}x_iy_i = \begin{cases}0 & x_i = y_i = 0\\ 1 & \text{其它}\end{cases}$.

这里 $0 \leqslant i \leqslant n-1$。

依定义 1, $n=2$ 时的 2^n 值逻辑真值表如下表所示。

2^2 值逻辑的真值表

y / x	$\neg x$	$x \vee y$			
		11	10	01	00
11	00	11	11	11	11
10	01	11	10	11	10
01	10	11	11	01	01
00	11	11	10	01	00

定义 2 设 $x, y \in S_{2n}$.

(1) 若 $x \wedge y = \neg(\neg x \vee \neg y)$, 则称符号∧为合取联结词;

(2) 若 $x \to y = \neg x \vee y$, 则称符号→为蕴涵联结词;

(3) 若 $x \leftrightarrow y = (x \to y) \wedge (y \to x)$, 则称符号↔为双向蕴涵联结词。

定理 1 设 $x, y \in S_{2n}$, $x = x_{n-1}x_{n-2}\cdots x_0$, $y = y_{n-1}y_{n-2}\cdots y_0$.

(1) 设 $x \wedge y = {}^{\wedge}x_{n-1}y_{n-1}{}^{\wedge}x_{n-2}y_{n-2}\cdots{}^{\wedge}x_0y_0$,则 ${}^{\wedge}x_iy_i = \begin{cases}1 & x_i = y_i = 1 \\ 0 & \text{其它}\end{cases}$;

(2) 设 $x \to y = {}^{\to}x_{n-1}y_{n-1}{}^{\to}x_{n-2}y_{n-2}\cdots{}^{\to}x_0y_0$,则 ${}^{\to}x_iy_i = \begin{cases}0 & x_i = 1, y_i = 0 \\ 1 & \text{其它}\end{cases}$;

(3) 设 $x \leftrightarrow y = {}^{\leftrightarrow}x_{n-1}y_{n-1}{}^{\leftrightarrow}x_{n-2}y_{n-2}\cdots{}^{\leftrightarrow}x_0y_0$,则 ${}^{\leftrightarrow}x_iy_i = \begin{cases}0 & x_i \neq y_i \\ 1 & x_i = y_i\end{cases}$.

这里 $0 \leqslant i \leqslant n-1$。

证明　选证 (2). 根据定义 2,因为 $x \to y = \neg x \vee y$, 所以

$x \to y = ({}^{\neg}x_{n-1}{}^{\neg}x_{n-2}\cdots{}^{\neg}x_0) \vee (y_{n-1}y_{n-2}\cdots y_0) = {}^{\vee\neg}x_{n-1}y_{n-1}{}^{\vee\neg}x_{n-2}y_{n-2}\cdots{}^{\vee\neg}x_0y_0$.

考虑 ${}^{\vee\neg}x_iy_i$, 根据定义 1,当 $x_i = 0$ 时, ${}^{\neg}x_i = 1$, ${}^{\vee\neg}x_iy_i = {}^{\vee}1y_i = 1$;

当 $x_i = 1$ 时, ${}^{\neg}x_i = 0$, ${}^{\wedge\neg}x_iy_i = {}^{\vee}0y_i = y_i$.

所以, 只在 $x_i = 1$ 和 $y_i = 0$ 时, ${}^{\vee\neg}x_iy_i = 0y_i$.

令 ${}^{\vee\neg} = {}^{\to}$,即得所需。　□

定理 2　$\neg\neg x = x$.

证明. 根据定义 1(1), 设 $x_i = 1$, 则 ${}^{\neg}x_i = 0$, ${}^{\neg\neg}x_i = {}^{\neg}0 = x_i$;

设 $x_i = 0$, 则 ${}^{\neg}x_i = 1$, ${}^{\neg\neg}x_i = {}^{\neg}1 = x_i$.

故 ${}^{\neg\neg}x_i = x_i$.　(1)

现设 $x = x_{n-1}x_{n-1}\cdots x_0$. 根据定义 1(1), 有

$\neg x = {}^{\neg}x_{n-1}{}^{\neg}x_{n-2}\cdots{}^{\neg}x_0$　(2)

根据式(2)、定义 1(1)和式(1), 有

$\neg\neg x = {}^{\neg\neg}x_{n-1}{}^{\neg\neg}x_{n-2}\cdots{}^{\neg\neg}x_0 = x_{n-1}x_{n-2}\cdots x_0 = x$.　□

引理 3　$x \vee y = \neg x \to y$.

引理 4　$x \wedge y = \neg(x \to \neg y)$.

引理 5　$x \leftrightarrow y = \neg((x \to y) \to \neg(y \to x))$.

定理 6　$\{\neg, \to\}$ 是联接词含量完全的。

该定理由引理 3、引理 4 和引理 5 即得。

定理 7　(1) $(\neg x \vee y) = \neg x \wedge \neg y$;(2) $\neg(x \wedge y) = \neg x \vee \neg y$.

定理 8　(1) $(x \vee y) \wedge z = (x \wedge z) \vee (y \wedge z)$;(2) $(x \wedge y) \vee z = (x \vee z) \wedge (y \vee z)$;(3) $x \to (y \to z) = (x \to y) \to (x \to z)$.

2　公式与赋值

命题是指可以用真值区分的陈述句,本文用 p,q,r 等表示;公式是指命题,或含有命题和联结词的表达式,用 A, B, C 等表示。

定义 3　设 2^nP 的合适公式集为 Γ, 真值集为 S_{2n}。2^nP 的赋值是指映射 v: $\Gamma \to S_{2n}$.
并且对任意的 A, $B \in \Gamma$, 有

(1) $v(\neg A) = \neg v(A) = 2^n - 1 - v(A)$;

(2) $v(A \vee B) = v(A) \vee v(B) = 0$ 当且仅当 $v(A) = v(B) = 0$.

定理 9　$v(A) \vee v(\neg A) = 2^n - 1$.

证明. 设 $v(A) = x_{n-1}x_{n-2}\cdots x_0$, 那么

$$\begin{aligned} v(A) \vee v(\neg A) &= (x_{n-1}x_{n-2}\cdots x_0) \vee ({}^{\neg}x_{n-1}{}^{\neg}x_{n-2}\cdots{}^{\neg}x_0) && \text{; 定义 3, 定义 1(1)} \\ &= {}^{\vee}x_{n-1}{}^{\neg}x_{n-1}{}^{\vee}x_{n-2}{}^{\neg}x_{n-2}\cdots{}^{\vee}x_0{}^{\neg}x_0 && \text{; 定义 1(2)} \\ &= \underbrace{11\cdots1}_{n\ \text{bits}} && \text{; 定义 1(2)} \\ &= 2^n - 1. \end{aligned}$$
□

定理 10 $v(A \to B) = 2^n - 1$ if $v(A) = v(B)$.

证明. Let $v(A) = v(B) = x_{n-1}x_{n-2}\cdots x_0$.

$v(A \to B) = v(\neg A \vee B)$; 定义 2(2)

$= v(\neg A) \vee v(B)$; 定义 3(2)

$= ({}^{\neg}x_{n-1}{}^{\neg}x_{n-2}\cdots{}^{\neg}x_0) \vee (x_{n-1}x_{n-1}\cdots x_0)$; 条件, 定义 1(1)

$= {}^{\vee\neg}x_{n-1}x_{n-1}{}^{\vee\neg}x_{n-2}x_{n-2}\cdots{}^{\vee\neg}x_0x_0$; 定义 1(2)

$= \underbrace{11\cdots1}_{n \text{ bits}}$; 定义 1(2)

$= 2^n - 1.$ □

引理 11 $v(A \to B) = v(B)$ if $v(A) = 2^n - 1$.

证明.

$v(A \to B) = v(\neg A \vee B)$; 定义 2(2)

$= v(\neg A) \vee v(B)$; 定义 3(2)

$= (2^n - 1 - v(A)) \vee v(B)$; 条件, 定义 3(1)

$= (00\cdots0) \vee v(B)$

$= v(B).$ □ ; 定义 1(2)

定理 12 若 $V(A) = 2^n - 1$, $V(A \to B) = 2^n - 1$, 则 $V(B) = 2^n - 1$.

证明. 反设 $V(B) \neq 2^n - 1$。由引理 11 知,当 $V(A) = 2^n - 1$ 时, $V(A \to B) = V(B)$,即 $V(A \to B) \neq 2^n - 1$,与条件矛盾. □

定理 13 $v(A \to (B \to A)) = 2^n - 1$.

定理 14 $v((\neg A \to B) \to ((\neg A \to \neg B) \to A)) = 2^n - 1$

定理 15 $v((A \to (B \to C)) \to ((A \to B) \to (A \to C))) = 2^n - 1$

定理 16 $v((\neg A \to \neg B) \to (B \to A)) = 2^n - 1$

3 应用示例

例 1 2^n 值逻辑在规范大数据模型及其对数据的基本操作方面也许能发挥作用。在要求高性能(High performance)、高效率存储访问(Huge Storage)、可伸缩性和高可用性(High Scalability and High Availability)的大数据环境(4H)[11],传统的关系数据库在 I/O 操作、表结构更改等方面产生了困难。而以 key - value 为数据模型的非关系型的数据库 NoSQL 较好地适应了 4H。key - value 是具有 key 和 value 两列的表, key 是查找数据地址的唯一关键字,value 是该数据实际存储的内容。这种数据库的存储结构不固定,每个元组可以根据需要确定 key - value 的数量。尽管它们能够实现大数据的快速查询,但不同公司的 NoSQL 产品的功能以及采用的关键技术都存在很大差别,所以难于实现开放性。可以考虑采用逻辑方法规范定义,并根据连接词的功能设计基本操作。

例如,为新建 NoSQL 中的操作,记 key 值的集合为 K, 2^n 值逻辑的真值集合为 S_{2n}, 并建立了映射

$$f: K \to S_{2n}.$$

现有 3 个 key - value 对 (k_1, 李晓), (k_2, 逻辑), (k_3, 4)分别表示姓名、课程和成绩, 这里 k_1, k_2, $k_3 \in K$. 若 $f(k_1) \wedge f(k_2) = f(k_3)$, 那么由 k_3 检索到的 4 就是李晓所修的逻辑课程的成绩。

注记 1 这个例子带来了一个新课题,就是如何设计 f, 使之与后续的逻辑操作相关,以满足应用的需要。

例 2 在大数据中,数据冗余是必须处理的问题。适当的冗余有益于系统的高可靠性和高性能,但“降冗”,即去除部分冗余数据,是必要的。先建立两个概念:本源数据和派生数据。本源数据是指不可替代的永恒数据。本源数据的特征是内容不可刷新,但它的集合可以扩充。

派生数据是指使用本源数据后所产生的数据。很清楚,要消除的冗余数据包含在派生数据中。

派生数据可能成为本源数据。然而,若派生数据是可以通过产生它的方法而复现,那么它是否有必要成为本源数据? 因此,需要找到派生数据进化为本源数据的条件。分类是一种定性分析。可以设想,

首先将所有的数据划分为 2^1 大类：本源数据和派生数据。然后根据特征，将本源数据不断分下去，最终得到 2^m+j 类，从而形成一个合理的本源数据层次结构。在找到进化条件的前提下，2^n 值逻辑就可能应用于对降冗算法的设计。

例如，设命题 p 表示“通过计算得到的派生数据”，q 表示“派生数据转换为本源数据”。假设已有进化条件，并根据它设逻辑值 11、10、01 和 00 分别表示真且符合进化条件、真但不合条件、假但符合条件、假且符合条件。这里所有的命题的赋值和逻辑运算结果的赋值都有相应的现实解释。例如，若 v(p)＝11 且 v(p→q)＝10，则表明：虽然派生数据转换为本源数据了，但这种转换不符合进化条件的规定。这可能是本源数据集中已有了该数据，或它可以简便复现。

注记 2　设 v(p)＝00，那么不论 q 取什么值，都有 v(p→q)＝11。这表明“如果前提是假，但不符合条件，那么什么结果都会产生”是对的，这陷入了混乱状态。因此，我们在应用中需要格外小心。

4　结语

我们构建了 2^nP 的语义框架，讨论了 2^n 值逻辑语义应用于 key－value 模型的有效性，提出降冗概念并初步研究了数据冗余条件逻辑语义和现实解释。2^nP 顺应了大数据时代，因此，它可能为建立大数据科学发挥基础作用。由于电子数字计算机采用二进制数和数字逻辑，我们相信 2^nP 将拓宽计算机科学、人工智能和信息技术等领域研究的逻辑基础。

此外，从语义看，2^nP 的语义与经典逻辑的命题系统（CP）的语义的主要关系如下：

2^n 值逻辑的位运算¬，∨，→，∧，↔与经典逻辑联结词¬，∨，→，∧，↔的运算完全相同。所以 CP 是 2^nP 的基础。

当 n(1 时，可以证明两个系统完全相同，例如，两个系统的联结词非和析取的真值可以分别表示为

$$\neg x = 1 - x,\ x_1 \vee x_2 = \max(x_1, x_2);$$

当 $n>1$ 时，两个系统的差异就显现出了。例如，在 2^nP 系统中，设 $x_1=01$，$x_2=10$，那么

$$x_1 \vee x_2 (11 \neq \max(x_1, x_2).$$

显然，2^nP 的语义包含了 CP 的语义。2^nP 是 CP 语义方面的多值扩展。

参考文献：

[1] Weinberg A. Impact of large－scale science on the United States [J]. Science, 1961, 134(3473): 161－164.

[2] Aronova E, S. Karen Baker, N. Oreskes. Big science and big data in biology: from the international geophysical year through the international biological program to the long term ecological research (LTER) network, 1957 － present [J]. Historical Studies in the Natural Science, 2010, 40(2): 183－224.

[3] 李国杰，程学旗. 大数据研究：未来科技及经济社会发展的重大战略领域[J]. 中国科学院院刊，2012，27(6)：647－657.

[4] Steve B, David K, Michael C, et al. Visually exploring gigabyte data sets in real time [J]. Communications of the ACM, 1999, 42(8): 82－90.

[5] Nancy R, Michael B. Focus issue: conquering the data mountain [J]. Science Signaling, 2011, 4(160): 2－3.

[6] Bell G, Hey T, Alex S. Beyond the data deluge[J]. Science, 2009, 323(5919): 1297－1298.

[7] Mattmann A. Computing: A vision for data science [J]. Nature, 2013, 493(7433): 473－475.

[8] Lukasiewicz J. On three－valued logic [J]. The Polish Review, 1968, 13(3): 43－44.

[9] Post E L. Introduction to a general theory of elementary propositions [J]. American Journal of Mathematics, 1921, 43(3): 163－185.

[10] Belnap B, JR. Entailment and relevance [J]. The Journal of Symbolic Logic, 1960, 25(2): 144－146.

[11] Rys M. Scalable SQL [J]. Communications of the ACM, 2011, 54(6): 48－53.

（责任校对　王小飞）