编码——隐匿在计算机软硬件背后的语言

A	1 444	J	- 100, 300, 500	S	***
В	una +++	K		Т	-
С	para e despira	Ī	2 April 7 G	U	a e inda
D	200 . 4. 4	M	लंका करन	ν	# 0 % PMM
E	•	N		W	.==
F		0		X	==
G	place adjug. a	P		Y	
Н	4444	Q		Z	
ſ	**	R			

以下是转化为比特的结果。

A	101100	J	101101101100	S	1010100
В	1101010100	K	110101100	Т	1100
С	11010116100	Ł.	1011010100	U	10101100
D	11010100	M	1101100	V	1010101100
E	100	N	110100	W	101101100
F	1010110100	О	1101101100	X	11010101100
G	110110100	P	10110110100	Y	110101101100
Н	101010100	Q	1101101011100	Z	11811818188
I	10180	R	10110100		

注意,这里所有编码都是以 1 开头,以两个 0 结尾。结尾处的两个 0 代表单词内各个字母之间的空格。单词之间的空格编码则是另外一对 0。因此,在莫尔斯码中"hi there"通常是这样表示的:

但是,当莫尔斯码采用比特来表示的时候,看起来就像 UPC 的一个截面。

用比特的形式表示布莱叶盲文比表示莫尔斯码容易得多。布莱叶盲文是 6 位编码。 其中的每一个字母都是由 6 个点组成的,这些点可能是凸起的,也可能是不凸起的。就 如同在第 3 章中所提到的那样,这些点通常用数字 1~6 编号。 1 O O 4 2 O O 5 3 O O 6

例如,单词 "code" 用布莱叶盲文可以这样表示:

..

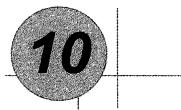
如果突起的点代表 1,平坦的点代表 0,则在布莱叶盲文中,每个符号都可以用 6 个比特的二进制代码来表示。单词"code"中的四个布莱叶字母符号可以简单地写成:

100100 101010 100110 100010

这里,最左边的一位对应编号为1的位置,最右边的一位对应编号为6的位置。

在本书后面的章节可以看到,比特可以表示文字、图片、声音、音乐、电影,也可以表示产品编码、胶片速度、影评结果、英国军队的入侵,以及心爱之人的意图。但是,从根本来说,比特是数字。在用比特表示其他信息的时候我们所要做的就是计算有多少种可能性。这决定了我们需要的比特位数,以便每种可能性都可以分配到一个编号。

比特在逻辑学中也很重要。逻辑学是哲学和数学的奇特融合,其主要目的就是确定某个陈述是真还是假。真和假同样可以表示为 1 和 0。



逻辑与开关

什么是真理? 亚里士多德认为真理是与逻辑相关的某种事物。《工具论》(可以追溯 到公元前 4 世纪)这本书收集了他的理论,这也是最早在逻辑上做出全面阐述的著作。 对于古希腊人来说,逻辑是在追求真理的过程中的一种分析方法,因此它被人们认为是 一种哲学形式。亚里士多德的逻辑学基础是三段论法。最著名的三段论法(但是这个三 段论却并没有出现在亚里士多德的著作中)是:

> 所有男人都必有一死; 苏格拉底是男人;

因此, 苏格拉底必有一死。

在三段论中,首先假定两个条件是正确的,然后再通过这两个条件推断出结论。

苏格拉底之死这种结论似乎太简单了,然而,三段论的形式其实是多种多样的。例如,以下两个由 19 世纪的数学家查尔斯・道奇森(也被人们称为刘易斯・卡罗尔)提出的条件:

所有哲人都是符合逻辑的;

一个没有逻辑的人总是非常顽固的。

上述条件的推论一点也不明显(结论是"一些顽固的人不是哲人。"注意,这里"一些"这个词的出现完全是意料之外的)。

在过去的两千多年里,数学家们都压制着亚里士多德的逻辑学,并试图用数学符号和算子限制它的发展。19世纪前期,唯一涉足这个领域的人就是莱布尼茨(1648-1716),他早年涉足了逻辑学,而后转移到了其他兴趣上(例如,与牛顿在同一时期,独立发明微积分)。

而后乔治・布尔出现了。

乔治·布尔于 1815 年出生于英格兰,降临到了这个对于他来说充满不公的世界。乔治的父亲是个鞋匠,母亲曾经做过女佣,因此,按常理在当时英国森严的等级制度下,他基本上不会做出什么有别于父辈们的事业。但是他凭借着勇于探究的精神以及父亲的帮助(乔治的父亲对科学、数学和文学都有浓厚的兴趣),乔治在年少的时候就接受了良好的教育,而这本应是上流社会中的男孩子们才能享有的特权。他学习了拉丁语、希腊语和数学。1849年,布尔凭借着他早期在数学研究上的论文成为了爱尔兰利克市皇后书院的数学系首席教授。



19 世纪中叶的一些数学家一直在研究逻辑的数学定义(其中最著名的是奥古斯特·德·摩根),但在理论上有着实际突破的却是布尔。首先是一部短篇著作《逻辑的数学分析——关于演绎推理的一篇随笔》(The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning,1847年),而后是一篇更长且更宏大的著作《思维规律的研究——逻辑与概率数学理论的基础》(An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities,1854年),通称《思维规律》(The Laws of Thought)。布尔死于1864年,他由于冒雨赶去上课而感染了肺炎,那一年,他年仅49岁。

布尔于 1854 年所发表著作,其书名就彰显了一种雄心壮志。因为理性的人总是通过逻辑去进行思考,如果我们找到了一种利用数学来描述逻辑的方法,也会找到一种用数学方法来描述大脑是如何工作的。当然,现在这种观点显得很是天真(然而在那个时代这确实是非常先进的思想)。

布尔发明了一种代数,这种代数看上去与传统代数非常相似,而且运算规则也非常类似。在传统的代数中,操作数(通常为字母)代表数字,算子(通常为"+"和"×")则用来指示这些数字之间如何运算。通常,我们应用传统代数来解决可能像这样的问题:安雅有 3 磅豆腐。贝蒂的豆腐是安雅的 2 倍。卡门的豆腐比贝蒂多 5 磅。迪尔得丽的豆腐是卡门的 3 倍。试问迪尔得丽有多少豆腐?

在解决这个问题时,我们首先要将文字叙述转化为数学语言,下面用四个字母分别 表示每个人所拥有的豆腐有多少磅:

$$A = 3$$

$$B = 2 \times A$$

$$C = B + 5$$

$$D = 3 \times C$$

将以上 4 个表达式带入同一式子中合并,最终可以得到一个加法和乘法混合运算的 式子:

$$D = 3 \times C$$

$$D = 3 \times (B + 5)$$

$$D = 3 \times ((2 \times A) + 5)$$

$$D = 3 \times ((2 \times 3) + 5)$$

$$D = 33$$

当进行传统代数运算的时候,我们会遵循一定的规则。这些规则在实际中可能非常根深蒂固,以至于我们不再认为它们是规则,甚至可能忘记它们的名字。但规则确实是任何形式数学运算的基础。

首先是加法和乘法的交换律(commutative)。意思就是,在运算符两边的操作数可以任意调换:

$$A + B = B + A$$
$$A \times B = B \times A$$

相反,减法和除法则是无法应用交换律的。

加法和乘法也遵循结合律 (associative), 如下:

$$A + (B + C) = (A + B) + C$$
$$A \times (B \times C) = (A \times B) \times C$$

最后,乘法遵循加法分配率 (distributive):

$$A \times (B + C) = (A \times B) + (A \times C)$$

传统代数的另一个特点就是,它是处理数字的,例如,豆腐的重量、鸭子的数量、 火车行驶的距离或家庭成员的年龄。布尔的天才之处就在于他把代数从数的概念中抽离 出来而使其更加抽象。在布尔代数(Boole's algebra,现在也这样叫)中,操作数不是数 字而是类(class)。简单说,一个类就是一个事物的群体,它后来也被称为集合(set)。

以猫为例。猫可以分为公猫与母猫。为了简便,用字母M代表公猫,字母F代表母猫。记住,这两个符号代表的不是数字。公猫与母猫的数量会随着小猫的出生和老猫的死去而发生变化。这两个字母分别表示两类猫——有特定特征的猫的群体。当我们提到公猫的时候,就可以使用M。

我们也可以用别的字母代表猫的颜色,例如:T可以代表褐色的猫,B可以代表黑猫,W代表白猫,O代表不在 T、B或 W集合中的其他颜色的猫。

仍以此为例,猫还可以分为已被绝育的和未被绝育的。我们用字母 N 来表示已被绝育的猫,字母 U 表示未被绝育的猫。

在传统的(数字的)代数中,符号"+"和"×"用来表示加法和乘法。在布尔代数中,也有"+"和"×"这样的符号,这很可能造成混淆。所有人都知道在传统代数中如何将数字相加或相乘,但是我们如何将类相加或相乘呢?

然而,在布尔代数中,并没有实际意义上的加或乘。而符号"+"和"×"表示的是完全不同的意义。

在布尔代数中,符号"+"表示两个集合的并集。两个集合的并集的意思就是指第一个集合中的所有元素与第二个类中所有元素的集合。例如,**B**+**W**表示的就是所有黑猫和白猫的集合。

在布尔代数中,符号"×"表示两个集合的交集。两个集合的交集就是指既在第一个集合中又在第二个集合中的所有元素的集合。例如, $F \times T$ 代表的是所有褐色母猫的集合。像传统代数一样,我们可以将 $F \times T$ 写为 $F \cdot T$,或者简写为 $F \cdot T$ (布尔代数中的首选形式)。可以将两个字母想象成两个形容词串联在一起:褐色的母猫。

为了避免在传统代数与布尔代数间混淆,有时用符号"U"和"∩"来代替"+"和 "×"。布尔对数学方面的革命性影响是让人们熟知的符号更加抽象,因此,我决定坚持他 的做法,而不是引入新的符号到他的代数中。

交换律、结合律和分配律都在布尔代数中同样成立。而且在布尔代数中,加法还可以来分配乘法,但在传统的代数中,这是不成立的:

$$W + (B \times F) = (W + B) \times (W + F)$$

白猫和黑色母猫的并集与如下两个集合的交集是一样的,这两个集合分别是,白猫和黑猫的并集,白猫和母猫的并集。虽然有点难理解,但是,这个结论是成立的。

布尔代数中还有另外两个符号是非常重要的。这两个符号看起来像数字,但是它们并不是真正意义上的数字,相对于数字而言,它们有不同的意义。在布尔代数中,符号 1表示"全集"——也就是我们所提到的所有事物。在前面的例子中,符号 1表示的就是"所有猫的集合"。因此:

$$M + F = 1$$

也就是说,公猫与母猫的并集是所有猫。同样的,褐色的猫、黑猫、白猫与其他颜色猫的并集也是所有猫的集合:

$$T+B+W+O=1$$

通过以下方法也可以得到所有猫的集合:

$$N + U = 1$$

符号1与减号连用可以表示在全集中排除一些事物,例如,

1- **M**

意思就是除去公猫的所有猫的集合。这个集合与母猫的集合是相等的:

1 - M = F

另一个要用到的符号就是 0, 在布尔代数中, 符号 0表示空集——不包含任何元素的集合。空集往往是两个互斥集合的交集, 例如, 母猫与公猫的交集:

$$\mathbf{F} \times \mathbf{M} = 0$$

注意,在布尔代数中,1和0有时也同在传统代数中的应用一样。例如,所有猫与母猫的交集是母猫的集合:

$$1 \times F = F$$

空集与母猫的交集还是空集:

$$0 \times \mathbf{F} = 0$$

空集与母猫的并集则是母猫:

$$0 + \boldsymbol{F} = \boldsymbol{F}$$

但是有时也会出现与传统代数相悖的结果。例如,所有猫与母猫的并集是所有猫的集合:

$$1 + F = 1$$

这个式子在传统代数中就是没有意义的。

由于 F 是母猫的集合,(1-F)是所有非母猫的猫的集合,因此这两个集合的并集是 1:

$$F + (1 - F) = 1$$

而这两个集合的交集为0:

$$\mathbf{F} \times (1 - \mathbf{F}) = 0$$

在历史上,这个公式代表了逻辑学上的一个重要概念:这一概念被称为矛盾律。矛盾律指出事物不可能既是它本身,同时又是它的对立面。

布尔代数中与传统代数形式上最大区别之处就是这样一个表达式:

$F \times F = F$

这个表达式很明确地表达了布尔代数的意义:母猫和母猫的交集依然是母猫。但是如果 F 代表的是数字,这个表达式就不会成立。布尔认为式子:

$$X^2 = X$$

就是将其代数同传统代数区别开的一条语句。另一个在传统代数中看起来比较有趣的式 子是:

$$F + F = F$$

母猫和母猫的并集依然是母猫。

布尔代数提供了一种解决亚里士多德三段论的数学方法。我们再看看最著名的三段论法中的前两句,但是现在不区分性别:

所有人都难逃一死; 苏格拉底是人。

我们用P表示所有人的集合,M表示必有一死的事物,S表示苏格拉底的集合。

"所有人都难逃一死"是什么意思? 意思就是说所有人的集合与必有一死的事物的交集是所有人:

$$P \times M = P$$

命题 $P \times M = M$ 是错误的,因为,在必有一死的事物中,还包含猫、狗和榆树等。

"苏格拉底是人"这句话的意思就是苏格拉底的集合(一个很小的集合)与所有人的集合(一个很大的集合)的交集是苏格拉底:

$$S \times P = S$$

根据第一个等式我们可以得到 $P = P \times M$,我们将这个等式带入到第二个等式中去:

$$S \times (P \times M) = S$$

根据结合律,可以写为:

$$(S \times P) \times M = S$$

并且我们已知 $S \times P = S$,因此上式可以简化为:

$$S \times M = S$$

到此,我们得出了结论。这个表达式告诉我们,苏格拉底的集合与必有一死的事物的集合的交集是 S,也就是说,苏格拉底必有一死。如果我们让 $S \times M = 0$,会得出苏格拉底不会死。如果 $S \times M = M$,结论就会是:只有苏格拉底会死,其他任何事物都是不死的!

应用布尔代数证明这个显而易见的事实(实际上在 2400 年前苏格拉底自己就证明了他必有一死)似乎有点小题大做,然而布尔代数同样可以用来确定某种事物是否遵循特定的标准。或许某天你走进了一家宠物商店,对店员说:"我想要一只公猫,已绝育的,白色或褐色都可以,或者一只母猫,也要是已绝育的,除了白色任何颜色都可以;或者,只要是黑猫就可以。"店员会对你说:"你想要的猫是在以下这样的集合里:

$$(\mathbf{M} \times \mathbf{N} \times (\mathbf{W} + \mathbf{T})) + (\mathbf{F} \times \mathbf{N} \times (1 - \mathbf{W})) + \mathbf{B}$$

对么?"你会说:"是的!正是!"

为了确定店员是正确的,你大概会去抛弃并集和交集的概念,并用 OR 和 AND 取而代之。这里将字母大写是因为,它们表示的不仅是字面意义,而表示布尔代数中的运算。做并集的时候,可以看做:第一个集合 OR 第二个集合。做交集的时候,可以看做:第一个集合 AND 第二个集合。除此之外,NOT 可以看做在 1 后面加一个减号。总的来说:

- ◈ 符号 "+"(之前作为并集的符号)现在可以用 OR 来表示。
- ♥ 符号 "×"(之前作为交集的符号)现在可以用 AND 来表示。
- 参 符号"1-"(之前意思是从全集中去掉某些元素)现在用 NOT 来表示。

因此,原表达式可以写为:

(M AND N AND (W OR T)) OR (F AND N AND (NOT W)) OR B

这样就非常接近你所说的话了。注意,这里是如何用括号表述清楚你的意图的。你想要的猫来自以下三个集合中的一个:

(M AND N AND (W OR T))

OR

(FAND NAND (NOT W))

OR

B

有了这个公式,店员就可以做一个布尔测试了。为了避免麻烦,我采用了一种略微有点不同的布尔代数形式——字母不仅仅代表集合。这里,字母可以用数字来赋值。我们只用数字 0 和 1。数字 1 代表 YES, True,即这只猫是符合这样的标准的。数字 0 表示NO, False,即这只猫不符合这种特定标准。

首先,店员拿出了一只未绝育的褐色公猫。以下是我期望得到的猫的表达式:

$$(\mathbf{M} \times \mathbf{N} \times (\mathbf{W} + \mathbf{T})) + (\mathbf{F} \times \mathbf{N} \times (1 - \mathbf{W})) + \mathbf{B}$$

以下是用0和1替换之后的式子:

$$(1 \times 0 \times (0+1)) + (0 \times 0 \times (1-0)) + 0$$

注意,只有M和T的值为1,因为这只猫是褐色的公猫。

现在我们要做的就是化简这个表达式。如果简化结果为 1,则这只猫就符合你的标准;如果简化结果为 0,那么这只猫就不符合。在化简的时候要切记,我们并不是真的在进行加和乘的运算,尽管通常我们可以当做是。符号"+"代表 OR,符号"×"代表 AND,在大多同类的规则中适用(有时在现代课本中也使用符号"△"和"∨"来表示 AND 和OR,而非用符号"×"和"+"。但是在这里符号"+"和"×"是具有特殊意义的)。

当符号 "×"代表 AND 时,有以下几种可能的结果:

$$0 \times 0 = 0$$

 $0 \times 1 = 0$

 $1 \times 0 = 0$

 $1 \times 1 = 1$

也就是说,只有当运算符 AND 左右两端都为 1 的时候,结果才为 1。这个运算与常规的乘法是完全相同的,该规则可以总结为如下这样一个表格,这与第 8 章中加法和乘

法表的形式类似。

AND	0	1
0	0	0
1	0	1

当符号"+"表示 OR 的时候,有以下几种可能的结果:

$$0 + 0 = 0$$

$$0 + 1 = 1$$

$$1 + 0 = 0$$

$$1 + 1 = 1$$

如果运算符 OR 左右两端有一个操作数为 1,则其运算结果就为 1。这个运算与常规的加法类似,但是在 1+1=1 这里例外。运算符 OR 可以概括为如下这样一个表格。

OR	0	1
0	0	1
1	1	1

我们应用这两个表格来计算一下原表达式的结果:

$$(1 \times 0 \times 1) + (0 \times 0 \times 1) + 0 = 0 + 0 + 0 = 0$$

结果为 0, 意味着 NO, False, 这只小猫不符合标准。

接着,店员拿出了一只已绝育的白色母猫。原始表达式是:

$$(\mathbf{M} \times \mathbf{N} \times (\mathbf{W} + \mathbf{T})) + (\mathbf{F} \times \mathbf{N} \times (1 - \mathbf{W})) + \mathbf{B}$$

将0和1带入得:

$$(0 \times 1 \times (1+0)) + (1 \times 1 \times (1-1)) + 0$$

化简得:

$$(0 \times 1 \times 1) + (1 \times 1 \times 0) + 0 = 0 + 0 + 0 = 0$$

因此, 第二只小猫也不符合标准。

然后,店员拿出了一只已绝育的灰色母猫(灰色属于其他颜色——非白色、黑色或褐

色)。表达式如下:

$$(0 \times 1 \times (0+0)) + (1 \times 1 \times (1-0)) + 0$$

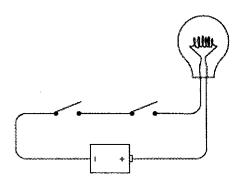
化简结果为:

$$(0 \times 1 \times 0) + (1 \times 1 \times 1) + 0 = 0 + 1 + 0 = 1$$

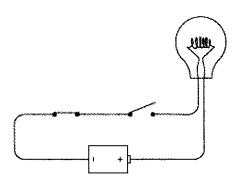
最后结果为 1, 意味着 Yes, True, 这个只小猫可以带回家(而且还是只漂亮的小猫!)。

当晚,在小猫蜷缩在你的腿上睡觉的时候,你突发奇想是否可以通过连通开关和灯泡的方法来确定某类猫咪是否符合你的标准(是的,你就是个奇怪的孩子)。你根本没发现,你将做出一个重要的概念上的突破。你要做的这些实验将布尔代数与电路相融合,并且可以设计和制造利用二进制进行计算的计算机。但是,不要让这些吓倒你。

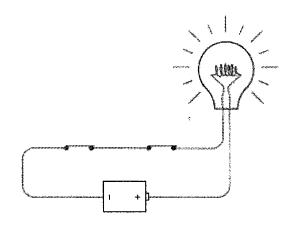
在实验的开始,将灯泡和电池正常连接起来,但是,你用了两个开关,而不是一个, 如下图所示。



像这样一个接着一个首尾相连的接线方式称做串联(series)。如果你闭合左端的开关,什么都不会发生。



如果让左边的开关保持断开,而闭合右边的开关,同样什么都不会发生。只有当左右两个开关都闭合时灯泡才会亮,如下图所示。



这里的关键词是"与"。当左右两个开关必须都是闭合的时候电流才流过回路。

这个电路是一个简单的逻辑演示。实际上,灯泡是在回答"两个开关是否都闭合?" 这样一个问题。这个电路的运转情况,我们可以总结为如下表格。

左开关	右开关	灯泡
断开	断开	不亮
断开	闭合	不亮
闭合	断开	不亮
闭合	闭合	亮

在之前的章节中,我们已经知道二进制数是如何表示信息的——而信息是无所不包的,无论是简单的数字还是罗杰·艾伯特²拇指的方向都是信息。我们可以说二进制 0 代表"艾伯特的拇指向下",二进制 1 代表"艾伯特的拇指向上。"一个开关有两个状态,因此可以代表二进制数。我们可以说 0 代表"开关断开",1 代表"开关闭合"。一个灯泡有两种状态,因此,它也可以用二进制数来表示。我们可以说 0 代表"灯泡不亮",1 代表"灯泡亮"。根据以上表述,我们可以将表格简化为如下形式。

² 美国影评家一译者注

左开关	右开关	灯泡
0	0	0
0	1	0
1	0	0
1	1	1

注意,如果将左边的开关和右边的开关调换位置,结果是一样的。我们不用分辨开 关哪个是哪个。所以,上面的表也可以写为和 AND 表或 OR 表类似的形式。

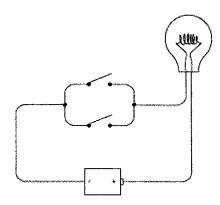
串联的开关	0	1
0	0	0
1	0	1

的确,这与 AND 表是一样的。

AND	0	1
0	0	0
1	0	1

这个简单的电路演示了布尔代数中的 AND 运算。

接下来,稍微改变一下开关连接方式,如下图所示。



这种连接方式叫做并联(parallel)。这种连接与上一种连接的不同之处就在于,闭合上面的开关,灯泡就会亮。