

# 因果知识的表示和推理<sup>\*</sup>)

林海 孙吉贵

(吉林大学计算机科学与技术学院 长春130012)

(吉林大学符号计算与知识工程教育部重点实验室 长春130012)

**摘要** 本文介绍了因果推理的两个主要应用:预测行为的间接结果,找出给定事实的真正原因,指出了用逻辑描述的因果关系在解决这两个问题中存在的不足。本文也简单地介绍了J. Pearl提出的因果推理方法,该方法的基本思想是把因果推理看成是一种“计算模式”,而这种把智能归结为“计算模式”的想法正是认知科学中的核心思想。本文最后比较了基于逻辑的方法和基于认知科学的方法实现智能的区别,指出了因果推理应该在认知科学的框架内得到解决。

**关键词** 因果推理,因果关系

## The Representation and Reasoning of Causal Knowledge

LIN Hai SUN Ji-Gui

(Computer Science and Technology Institute of Jilin University, ChangChun 130012)

(Key Laboratory of Symbolic Computation and Knowledge Engineer of Ministry of Education, ChangChun 130012)

**Abstract** This paper presents two main applications of causal reasoning: to predict indirect facts of actions and to find the actual cause of some given facts. We point out that logic is not adequate for solving these two problems. We also present J. Pearl's method in dealing with causal reasoning, which treats causation as a computational schema. This idea of treating intelligence as computational schema is the basic idea of cognitive science. As last, we contrast methods which implement intelligence based on logic and those based cognitive science, point out that causal reasoning should be performed in the framework of cognitive science.

**Keywords** Causal reasoning, Causation

### 1. 引言

我们生活的世界中充满了因果关系。知道了事物或者行为之间的因果关系显然对于指导我们的行为有重要的作用。假如人们知道乱砍乱伐会导致生态平衡遭到破坏,就不会再去乱砍乱伐;知道了不规律的饮食习惯会影响健康,人们就会注意养成好的饮食习惯。

然而因果关系的发现并不容易,因为它不是简单的时间上的接续关系:每天公鸡打鸣,然后天亮,但并不意味着公鸡打鸣引起太阳升起,否则人们简单地让公鸡早点打鸣就可以延长白天的时间。这是一个最简单的例子,面对这个复杂的世界,人们怎样从中找到因果关系?因果关系到底怎样表示?

最早探索因果关系的是 David Hume(1771-1776),他指出“*We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by object similar to the second, Or, in other words, where, if the first object had not been, the second had never existed.*”这种用判定因果关系的方法后来又被进一步发展成了反事实(conterfactual)的方法。然而,反事实的方法在实际中也遇到了一些困难。

1999年J. Pearl<sup>[2]</sup>提出了一种新的解决因果推理问题的方法,他是这样描述这种方法的主要思想的:“*The ruling conception will be to treat causation as a computational schema devised to identify the invariant relationships in the*

*enviroment, so as to facilitate reliable prediction of the effect of actions.*”而这种智能看成是一种“计算模式”的想法正是认知科学(cognitive science)中的核心思想。

本文首先列举了两个有关因果推理的例子,然后指出了用逻辑描述的因果关系的不足之处。接下来简单地介绍了J. Pearl提出的方法。最后从基本思想出发,比较了基于逻辑的方法和基于认知科学的方法实现人工智能的区别,指出了因果推理应该在认知科学的框架内得到更好的解决。

### 2. 两个有关因果推理的例子

我们首先从两个具体的例子出发来看一下用逻辑的方法描述因果关系存在的问题。

**例1<sup>[1]</sup>** 有一个皮箱,当它上面的两把锁都打开时皮箱也处于打开的状态。问题是:当前有一把锁处于打开的状态,另一把锁未被打开。如果现在打开另一把锁,问该皮箱的状态。

显然根据我们的直观:两把锁都打开会引起(cause)皮箱打开,因此皮箱应该也被打开。但如果把这个简单的因果关系用逻辑描述就会得出跟直观相反的结论:

$$\text{up}(L_1, s) \wedge \text{up}(L_2, s) \rightarrow \text{open}(s) \quad (1)$$

其中, $L_1$ 、 $L_2$ 分别代表两把锁, $s$ 代表当前状态。把(1)作为约束条件,满足(1)的状态都是该问题的答案。这样,会出现两个答案:1. 两把锁都被打开,皮箱也被打开;2. 原来开着的锁现在被关上,皮箱仍然处于关闭状态。后一个结果显然跟直观相反。

<sup>\*</sup>国家自然科学基金(60073039, 60273080)和吉林省科技发展计划(20020306)资助课题。

作者的解释是:该问题中只包含一种机制(mechanism),用(1)描述该问题隐含地把和(1)等价的  $up(L_1, s) \wedge \neg open(s) \rightarrow up(L_2, s)$  所代表的另外一种机制人为地加进来了,造成出现第二个答案。简单地说,  $A \rightarrow B$  等价于  $\neg B \rightarrow \neg A$ , 但因果关系是单向的,  $A \text{ cause } B$  不等价于  $\neg B \text{ cause } \neg A$ 。

例2<sup>[2]</sup> 如图1所示的电路图中。

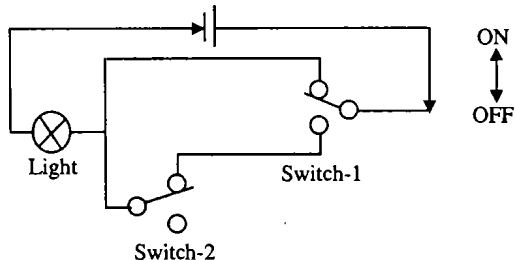


图1

显然,灯L亮当且仅当至少有一个开关处于on状态,如果把on状态计为1,则下面的关系成立: $L = SW_1 \vee SW_2$ 。SW<sub>1</sub>和SW<sub>2</sub>处于等同的地位,根据Mackie的INUS条件(将在第2部分介绍),SW<sub>1</sub>和SW<sub>2</sub>都是灯L亮的原因。但我们很容易看出只有SW<sub>1</sub>才是L亮的真正原因(电流根本不经过SW<sub>2</sub>)。

出现以上问题的原因是:逻辑公式只是从功能上(或者说结果上)描述变量之间的关系,而忽略了“内部结构”,而事实上我们正是通过仔细观察电路的内部结构才判断出真正原因的。

### 3. Mackie的INUS-condition:在逻辑框架内描述的因果关系

Mackie在1965年就试图在逻辑框架内描述因果关系。根据Mackie的论述,一个事件的原因既不是这个事件的充分条件也不是它的必要条件。而是“an *insufficient but necessary part of a condition which itself is unnecessary but sufficient for the result*”。这就是著名的INUS(四个斜体单词的缩写)条件。1988年R. W. Wright提出的NESS条件(Necessary Element of Sufficient Set)可以看作INUS条件的具体解释,这里我们主要讨论INUS条件。

例如,假设我划着一根火柴,根据常识有理由说我划火柴是火柴燃烧的原因。但它既不是充分条件也不是必要条件:如果空气中没有氧气或者火柴是湿的,即使划火柴,火柴也不会燃烧,因此它不是充分条件;另外可能有很多其它的原因能使火柴燃烧,不一定非需要划火柴,因此它也不是必要条件。而划火柴却是火柴燃烧的INUS条件: $B = (A \wedge X) \vee Y$ ,其中B表示火柴燃烧,A表示划火柴,X表示其它条件都正好合适(例如,空气中有充足的氧气,火柴不是湿的),Y表示其它能引起火柴燃烧的原因。

Mackie的INUS条件遇到了很多困难和批评。文[3]中指出:任意的事件C都可能成为一个事件P的INUS条件。假设  $A \wedge B$  是P的极小充分条件(简单地说就是去掉任何一个合取式后,剩余的部分都不再是充分条件,文[3]中有确切的定义),  $C \wedge (\neg C \vee A) \wedge B$  也是P的极小充分条件(除非一些特殊情况,例如  $\neg C \vee A = A$ )。也就是说,任何的事件C都有可能作为P的INUS条件,进而作为P的原因。

文[3]中对INUS的另外一点批评是它对表示方式有很大的依赖性:同样一个事件用不同的但彼此等价的逻辑公式

表示出来,INUS条件的集合可能不相同。例如,  $A \vee B$ ,  $A \vee (\neg A \wedge B)$  和  $B \vee (\neg B \wedge A)$  是等价的,  $\neg A$  在  $A \vee (\neg A \wedge B)$  的INUS条件之集中,  $\neg B$  在  $B \vee (\neg B \wedge A)$  的INUS条件之集中,而它们都不在  $A \vee B$  的INUS条件之集中。而直观告诉我们:同样一个事件的原因应该是一定的,不应该因表示方式的不同而发生变化。可能有人会提出应该用最简的形式来描述事件,比如这个例子中的  $A \vee B$ 。但是,就一般的公式来说,它的最简形式未必唯一。更重要的是也不能说用最简的形式描述就一定适合,例如,第1部分中的第2个例子就是因为用的是最简的描述形式才造成和直观不一致的结论。

### 4. 背景

对因果知识的描述和推理主要要解决两个问题:一个是对行为(action)进行推理,预测其结果(尤其是间接结果),例如第一部分的例1。第二个就是找出给定事实的真正原因,例如第一部分的例2。下面对它们分别进行介绍。

根据文[4]中的介绍,对行为的推理开始于John McCarthy 1959年的论文“Programs with Commonsense”,这篇论文中指出了对行为的推理在常识推理中占有非常重要的地位。后来在1963年McCarthy提出了Situation Calculus来描述行为域(action domain)。大概在80年代,人们发现经典的逻辑是单调的,不足以描述常识推理中的问题,这期间提出了很多非单调的逻辑。例如,McCarthy的circumscription, Reiter的default logic和Moore的autoepistemic logic。20世纪90年代,人们在发现解决ramification问题,即预测一个行为的间接结果(indirect effect)的问题中,用状态约束(state constraint)描述背景知识是不够的,因果知识(causal knowledge)取代了状态约束,被用来描述背景知识。相关的工作主要有文[4~10]。

值得注意的是,这些工作虽然用到了因果知识但都没有明确地给出因果关系的定义。例如,文[4]中提到:只要给出一个事件就发生的条件就已经足够了,假设有这样三种类型的知识:

- 1 The fact A causes the fact B.
- 2 Necessarily, if A then the fact B is caused.
- 3 Necessarily, if A then B.

这三种类型的知识强弱程度是递减的。1明确表示出了A和B的因果关系,2表示只要A成立,一定存在这样的事件(不一定是A本身,还可能有其它的原因),它能够引起B也成立。3中不包含任何因果关系。在解决ramification问题时,我们只需要第二种知识就够了。

有关因果推理中的第二个问题是找出某个事件发生的(真正)原因。对这个问题的研究比较早,除了上面提到的INUS之外,很经典的也是很直观的方法是所谓的“反事实方法(counterfactual approach)<sup>[11]</sup>”。即:A是B的原因,如果A不发生B也不会发生。但事实上A已经发生了,“假设A不发生”跟事实相反,所以叫反事实的方法。例如上面划火柴的问题,说划火柴是火柴燃烧的原因是因为(根据反事实的方法):如果不划火柴,火柴自己是不会燃烧的。反事实的方法在实际中也遇到了困难,考虑下面的问题:

例3 一个小球高速飞向一块玻璃,在小球的飞行路线上(小球和玻璃之间)有两个障碍物。每一个障碍物都足以阻止小球继续飞行。结果当然是小球的飞行被第一个障碍物阻止,使得玻璃没有被小球撞碎。问题是什么是玻璃没有被小球撞

碎的原因。

直观上应该是第一个障碍物。但是如果按照反事实的方法:假设没有第一个障碍物,小球也会被第二个障碍物阻止,玻璃仍然不会被撞碎,所以第一个障碍物不是玻璃没被撞碎的原因。这种情况通常被叫做 overdetermined,直观上说就是可能有一些干扰因素,它们的存在并没有改变整个因果过程,只是使得问题变得复杂,比如这个例子当中的第二个障碍物。如果去掉这些干扰因素,比如拿掉第二个障碍物,问题才会变得清楚。

### 5. 以 J. Pearl 的方法为代表的近期相关工作

J. Pearl 在文[2]中比较好地解决了上述两个问题,该方法的核心思想是使用一些“modifiable structural equations”。所谓“structural equation”是指方程和它的结构有关,方程两边变量的地位不同,这种思想源自于早期的经济学。例如,  $consume = income/2 + \epsilon$ , 收入和消费之间有一个因果关系,如果收入提高了,带来的间接结果是消费也提高了,但反过来不成立。因果关系中含有单向的意思,即,如果 A 是 B 的原因那么 A 的改变自然地会引起 B 的改变,反之不然。所谓“modifiable”是指其中的任何一个方程都可能被其它的方程所替代,如下面的定义 5.2,这种替代可能是由外界的干涉(intervention)引起的。

**定义 5.1** 一个因果模型(causal model)是一个三元组:

$M = \langle U, V, F \rangle$ , 其中

(1)  $U$  是由模型外的因素决定的背景变量(外部变量)的集合。

(2)  $V$  是由模型内的变量决定的内部变量的集合,即由  $U \cup V$  决定的变量的集合。

(3)  $F = \{F_1, F_2, \dots, F_n\}$  是函数的集合,其中  $F_i$  是从  $U \cup (V \setminus V_i)$  到  $V_i$  的映射。

每个因果模型都对应着一个有向图,图中的节点相当于  $V$  中的变量,边由决定某个变量的变量指向被决定的变量,把这样的图叫做因果图(causal diagram)。这里没有给出一种如何确定内部变量和外部变量的具体方法,因为我们每个人看问题的角度可能不同,往往把我们关心的那一部分事物划分到内部变量中,不关心的那一部分划分到外部变量中。

**定义 5.2**(子模型, submodel) 令  $M$  为一因果模型,  $X \subseteq V$ ,  $M$  的子模型  $M_x$  是因果模型:  $M_x = \langle U, V, F_x \rangle$ , 其中

$F_x = \{F_i | V_i \notin X\} \cup \{X=x\}$

也就是,把  $F$  中所有对应于  $X$  中的变量的函数删除,用  $X=x$  取代。

如果我们令  $do(X=x)$  表示行为(action),则有下面的定义:

**定义 5.3**(行为的结果, Effect of action) 令  $M$  为一因果模型,  $X \subseteq V$ , 行为  $do(X=x)$  作用在  $M$  上的结果就是子模型  $M_x$ 。

**定义 5.4**(一个因果公式, causal formula) 是形如  $X=x$  的表达式的布尔组合。 $(M, u^*) \models \phi$  表示公式  $\phi$  在  $M$  中为真。 $(M, u^*) \models [Y^* = y^*] \phi$  表示公式  $\phi$  在子模型  $M_{Y^* = y^*}$  中为真(带“\*”的表示向量,下同)。

**定义 5.5**  $X^* = x^*$  在  $(M, u^*)$  中是  $\phi$  的真正原因,如果以下三个条件成立:

(1)  $(M, u^*) \models (X^* = x^*) \wedge \phi$  (即,原因和结果都要在现实世界中成立)。

(2) 存在  $V$  的一个划分  $(Z^*, W^*)$ , 使得  $X^* \subseteq Z^*$ , 并且如果  $(M, u^*) \models Z^* = z^*$ , 那么

a.  $(M, u^*) \models [X^* = x^*, W^* = w^*] \rightarrow \phi$ . (也就是说,把  $(X^*, W^*)$  从  $(x^*, w^*)$  改变为  $(x', w')$  会使  $\phi$  由真变为假。)

b.  $(M, u^*) \models [X^* = x^*, W^* = w^*] (\phi \wedge (Z^* = z^*))$ . (也就是说,只要  $X^*$  保持现有的  $x^*$  不变,改变  $W^*$  不会影响  $Z^*$  中的任何变量。)

(3)  $X^*$  是最小的,即,没有  $X^*$  的子集满足(1)和(2)。

一般把上面的  $Z^*$  叫做“active causal process”,实际上,  $Z^*$  的所有变量都在  $X^*$  中某个变量到  $\phi$  中某个变量的路径上,文[12]中对此有证明。定义 5.5 中的(1)和(3)都很好理解,(2)实际上是为了处理例 3 中的情况,  $W$  相当于一些干扰因素,我们允许改变  $W$  的值(见 a),前提是改变  $W$  的值对“active causal process”中的变量没有影响(见 b)。

至此,第 3 部分中提到的两个问题,“预测行为结果”的问题和“找给定事实原因”的问题,分别由定义 5.3 和定义 5.5 给出了回答。文[13]中对这种方法的复杂性进行了分析。下面结合例 1 看一下 J. Pearl 的方法在实际中怎样应用:

我们定义 3 个内部变量,  $suitcase\_open, L_1\_open, L_2\_open$ , 它们都是取 1 表示打开,取 0 表示关闭,定义 2 个外部变量,  $u_1$  和  $u_2$ , 分别表示决定  $L_1$  和  $L_2$  状态的因素(由于是外部变量,我们不去关心它们具体是什么),它们为 1 表示能引起相应锁为“开”状态的环境。在本例子中  $u_1=1, u_2=0$ 。函数  $F$  的集合由下面 3 个方程组成:

$$suitcase\_open = L_1\_open \wedge L_2\_open \quad (1)$$

$$L_1\_open = u_1 \quad (2)$$

$$L_2\_open = u_2 \quad (3)$$

把锁  $L_2$  的开的行为用  $do(L_2\_open = 1)$  表示。这样,在  $M_{L_2\_open=1}$  中(3)需要被  $L_2\_open = 1$  替换掉。得到新的函数的集合:

$$suitcase\_open = L_1\_open \wedge L_2\_open \quad (1')$$

$$L_1\_open = u_1 \quad (2')$$

$$L_2\_open = 1 \quad (3')$$

这样  $suitcase\_open = L_1\_open \wedge L_2\_open = 1 \wedge 1 = 1$ , 即皮箱也被打开。

### 6. 进一步的思考

#### 6.1 cognitive science vs. logical approach

实现人工智能是计算机科学追求的重要目标之一。实现人工智能的途径有很多,认知科学和逻辑的方法是其中主要的两种途径。逻辑的方法是我们比较熟悉的,下面简单介绍一下认知科学。认知科学其实并不是离我们很遥远的,神经网络其实就属于认知科学的范畴,认知科学的出发点其实也很简单。

认知科学认为:人脑之所以能实现智能是因为人脑中含有这样一个层次,在这个层次中含有一些内部表示(mental representation)和建立在这些内部表示之上的计算过程。这跟计算机中的数据结构和建立在数据结构之上的算法很相似。在这个层次上二者找到共同点,也就是实现人工智能的切入点。至于这个层次之下,计算机中是一些硬件和软件实现的计算过程,人脑中可能是一些物理的、化学的过程,也可能是其它的,我们不用去管它。认知科学的中心任务就是:在计算机

上寻找合适的数据结构和算法来对应于人脑中的内部表示和计算过程,从而实现智能。

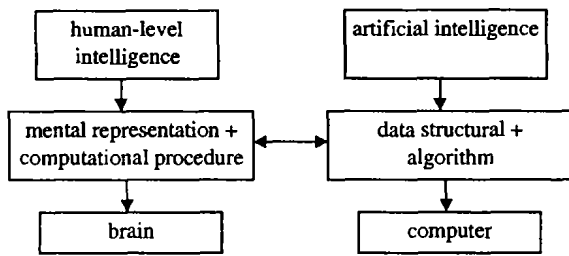


图1 认知科学中智能的实现

认知科学和逻辑方法实现人工智能的出发点是完全不同的,这一点 McCarthy 在“Artificial Intelligence, Logic and Formalizing Common Sense”中就有过论述。简单地说,用逻辑的方法实现智能是基于对常识世界(common-sense world)的理解,而认知科学实现的智能是基于对人脑的理解,试图弄清“黑匣子”内部真正的表示和推理。

1971年, John McCarthy 图灵奖演讲的题目是“Generality in Artificial Intelligence”,他希望一般的人工智能问题都能在逻辑框架内得到解决。但是,这要有两个前提条件:第一,所有的知识都能在逻辑框架内得到表示。第二,在逻辑框架内的推理要足够高效。遗憾的是,这两个前提条件都很难成立。一般的推理问题是 NP 问题,效率比较低;有关第一个前提条件,美国的学术界爆发过一场所谓的“civil war”。逻辑学派(logicists)一方以 Stanford 为代表,认知学派(proceduralists)一方以 MIT 为代表。认知学派认为整个逻辑系统缺乏专门的领域知识(domain-specific knowledge),过于静态(static)和刻板(rigid),很难想象动态的、丰富的世界能够完全都用逻辑刻画出来。引用当时 Papert 的话说就是:“As long as your methods are supposed to be good for proving anything, they're not likely to be good for proving anything.”

## 6.2 因果推理应该在认知科学框架内得到解决

如前所述,用逻辑描述的因果知识虽然能解决 ramification 问题,但却没有明确地回答“因果关系是什么”,“A 是 B 的原因究竟意味着什么”这样的问题。也就是说,由于逻辑中对因果知识的描述并不完全,导致它只能解决一部分因果推理问题。到目前为止, J. Pearl 给出的因果关系的定义是最合理的,也是形式化最好的。这种方法其实就是假定在人脑的某一层次上有这样一个 causal diagram(如前所述),人们在预测一个行为的间接结果和确定一事件的原因时实际的思维过程就像定义 5.3 和 5.5 所描述的那样。

正如前面介绍的,逻辑学派是基于对常识世界的理解,而认知学派是基于对人脑,对思维的理解。“因果关系是什么”的问题实际上就等于“人是怎么理解因果关系的”,换句话说就

是“人在面对因果关系的时候是怎样进行思维的”,这样的问题应该建立在对人类思维正确理解的基础上才能得到比较合理的答案,单从逻辑出发进行描述很可能忽视了因果推理的特殊性。

**结语** 本文主要讨论了逻辑在描述因果推理中的不足,而与之相对应的认知科学比较好地解决了因果推理中的一系列问题。逻辑和认知科学在实现智能上的主要区别是二者的出发点不同:逻辑的方法是基于人对常识世界的理解,而认知科学的方法是基于人对人脑的理解。但从逻辑的角度描述因果关系不能很好地解决因果推理中的一些问题,因而是不全面的。

## 参考文献

- 1 Lin F. Embracing causality in specifying the indirect effects of actions. In: Proc. IJCAI-95, 1995
- 2 Pearl J. Reasoning with Cause and Effect. In: Proc. of the Intl. Joint Conf. on Artificial Intelligence (IJCAI), 1999. 1437~1449
- 3 Kim J. Causes and Events; Mackie on Causation. in Causation and Conditionals, ed. Ernest Sosa, Oxford University Press, 1975. 48~62
- 4 McCain N. Causality in Commonsense Reasoning About Actions: [PhD thesis]. The University of Texas at Austin, 1997
- 5 Turner H. A logic of universal causation. Artificial Intelligence, 1999. 113: 87~123
- 6 Geffner H. Causality, constraints and the indirect effects of actions. In: Proc. IJCAI'97, Nagoya, Japan, 1997. 550~560
- 7 McCain N, Truner H. A causal theory of ramifications and quantifications. In: Proc. IJCAI-95, 1995
- 8 McCain N, Turner H. Causal theories of action and change. In: Proc. AAAI-97, 1997
- 9 Turner H. Representing Actions in Default Logic: A Situation Calculus Approach. In: Proc. Common Sense'96, 1996
- 10 Gelfond M, Lifschitz V. Representing action and change by logic programs. Journal of Logic Programming, 1993, 17: 301~322
- 11 Lewis D. Counterfactuals. Harvard University Press, Cambridge, MA, 1973
- 12 Halpern J Y, Pearl J. Causes and explanations: A structural-model approach. In IJCAI-2001, pages 27~34
- 13 Eiter T, Lukasiewicz T. Complexity results for structure-based causality. In IJCAI-2001, pages 35~40
- 14 Pearl J. On the definition of actual cause: [Technical Reprint R-259]. Department of Computer Science, University of California, Los Angeles, CA, 1998
- 15 Wright R W. Causation, responsibility, risk, probability, naked statistics, and proof: Pruning the bramble bush by clarifying the concepts. Iowa Law Review, 1988, 73: 1001~1077