

# 解 释 归 约

## ——解决多种解释问题的一种途径

张少平 陈火旺 (国防科技大学计算机科学系)

### 摘 要

在基于解释的学习中,主要问题是因不完善域论产生多种解释。本文提出了一种解释归约方法,能部分地解决多种解释问题。试验设计是解释归约方法的关键,本文讨论了它的理论和实现方法。

基于解释学习是机器学习领域中人们比较关注的学习方法,我们可以把该方法概括成下面两步:

1)构造解释来说明为什么给定的例子是目标概念的实例。

2)对解释进行一般化拓广,从而获得目标概念的学习描述。

构造解释是一个与域论紧密相关的构造过程,并且要求域论是完善的。然而实际应用中,这样的完善域论难得,因而基于解释学习方法的成功在很大程度上取决于能够在不完善域论上构造解释。

在构造解释过程中,由于不完善域论,为了说明给定的例子是目标概念的实例,将导致一组不相容的解释。标准的基于解释学习系统不能处理这种多种解释问题。多种解释问题是由不完善域论所产生的主要问题,其原因大致有以下三个方面:

(1)不完全的域论:由于知识不充分,在构造解释时必须作一些假设,每种假设都将产生一种解释。

(2)难处理的域论:在构造解释时,为了使问题易于处理,需作一些近似假设,从而产生了多种解释。

(3)不正确的域论:多种解释起源于不正确的域论。

让我们举一个机器人观察绳子的多种解释的简单例子。因为人们不可能完全地规范出世界的状态,故而机器人只能在不完全的信息状态下进行工作。当机器人观察到绳子变长这一事实后,它可能会给出下面三种不同解释:

- 1)绳子受热膨胀
- 2)绳子随时间生长
- 3)绳子是弹性的且正在受力

人们解决多种解释的方法之一是与环境交互,交互的结果是使用解释归约方法从环境中获取新的数据,从而排除一些解释。本文提出的这种方法能够部分地解决多种解释问题,并对解释归约方法的中心问题——试验设计进行了讨论。

### 一、解释归约

解释归约是一种解决多种解释问题的方法,它能够从一组候选的解释中找到一种与现实相符的解释,图1说明了解释归约系统的主要组件,解释归约过程包括下列三步:

#### 1)识别假设

该步骤要识别出支持解释的各种假设。

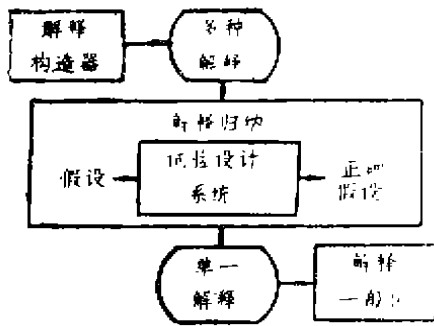


图1 解释归约系统图示

解释构造器可能对现实世界状态作出多种假设且基于每种假设构造解释。假如在上述机器人观察绳子的例子中，机器人可能作出绳子是弹性的、绳子受热及绳子生长三种假设。

### 2) 试验设计

试验设计是指通过测试假设的效果来直接或间接地验证假设。通过试验可以从现实世界中获取新的信息，后面我们将详细地描述试验设计过程。

### 3) 排除解释

根据试验所获取的信息，我们可以排除一些假设及基于这些假设所建立的解释。如果某假设的试验结果与期望结果不一致，我们就排除该种假设。

进行试验设计，需要有一定的策略支持。一般有三种策略来排除一些假设。

[a]求精 对于限定的论域，通常存在一些易于观察和测量的属性。在求精过程中，我们就选择这样的属性。如果某假设所推测的属性值与我们观察、测量的结果不一致，就可以立即排除该种假设。然而，该策略并不能保证所设计的试验一定能排除某种假设，例如，对于机器人观察绳子的例子，若我们选择易观察的绳子颜色属性，但因没有任何假设推测颜色的变化，故而试验不会排除任何假设。

[b]鉴别 在鉴别过程中，我们仅选择能够帮助排除假设的属性。我们说两个属性值是可区分的，是指存在某一测量能区分这两

个值。判别式是满足下面二条标准的属性。

- 存在一组假设，能够推测出该属性的不同属性值。

- 这些不同的属性值应当是可区分的。

如果判别式的属性值能够被划分成一组可区分且不相容的集合，并且对于该集合族中任何一个集合，都存在假设来推测此集合中的元素，则鉴别试验即可保证排除假设。显然鉴别是一种有效的策略，但代价较高。

[c]转换 对于某种试验设计脚本，无论我们怎样观察、测量，都无法识别出正确的假设，这时，脚本转换就成为了一种有效的试验设计方法。脚本转换要求以一种良定的方式改变初始脚本，例如，替换脚本构件，重新组织脚本构件等等，求精和鉴别技术又可用于新的脚本，从而排除假设。创建新脚本具有以下三个优点：

- 根据假设推测的不可区分的属性值，在新的脚本下，变成可区分的。

- 先前不可观察或不可测量的属性值，在新的脚本下，变成可观察或可测量的值。

- 有的属性虽然在旧脚本下可观察，其值可测量，但在新的脚本下，可观察属性会出现得快一点，或出现较多的可观察行为。

## 二、试验推理机及域论表示方法

试验推理机是基于上面描述的试验模型发展起来的，试验推理机的输入包括：

- 一组假设。
- 推理器 它能够接收假设和试验脚本，然后输出一组推测结果。
- 论域相关的知识 它由下述两种域知识组成：

a) 一组谓词 它们描述论域中能够被观察和测量到的属性。

b) 一组脚本转换操作 这些操作能够根据给定的脚本构造新的脚本。

试验推理机使用求精、鉴别和转换技术来设计试验。在这里转换实质上是一种规划问题 (Planning Problem)，初始状态是给定的脚本，目标状态是转换脚本，而规划是

产生目标状态的一系列转换。

基于解释学习方法是与域论紧密相关的，为了说明解释归约方法，我们对论域进行定性描述，该种域论表示是基于过程的，每一过程描述由五部分组成：

1)对象；2)前置条件；3)属性要求；4)后置条件；5)效果。

对象是指参与过程的个体，前置条件和属性要求规范了为激活过程所需满足的条件，而后置条件和效果则说明过程产生后的影响。

例如，对于化学分解这一过程，我们可

以进行如图2和图3的描述。

```

Decomposition:
  Objects:
    ? Substance? Product1? Product2
  Preconditions:
    (decomposes? Substance)
  Quantity Conditions:
  Relations:
    rate Q +amt(? Substance)
    rate Q -amt(? Product1)
    rate Q -amt(? Product2)
  Influences
    I+[amt(? Product1), rate]
    I+[amt(? Product2), rate]
    I-[amt(? Substance), rate]
    
```

图2 化学分解的定性描述

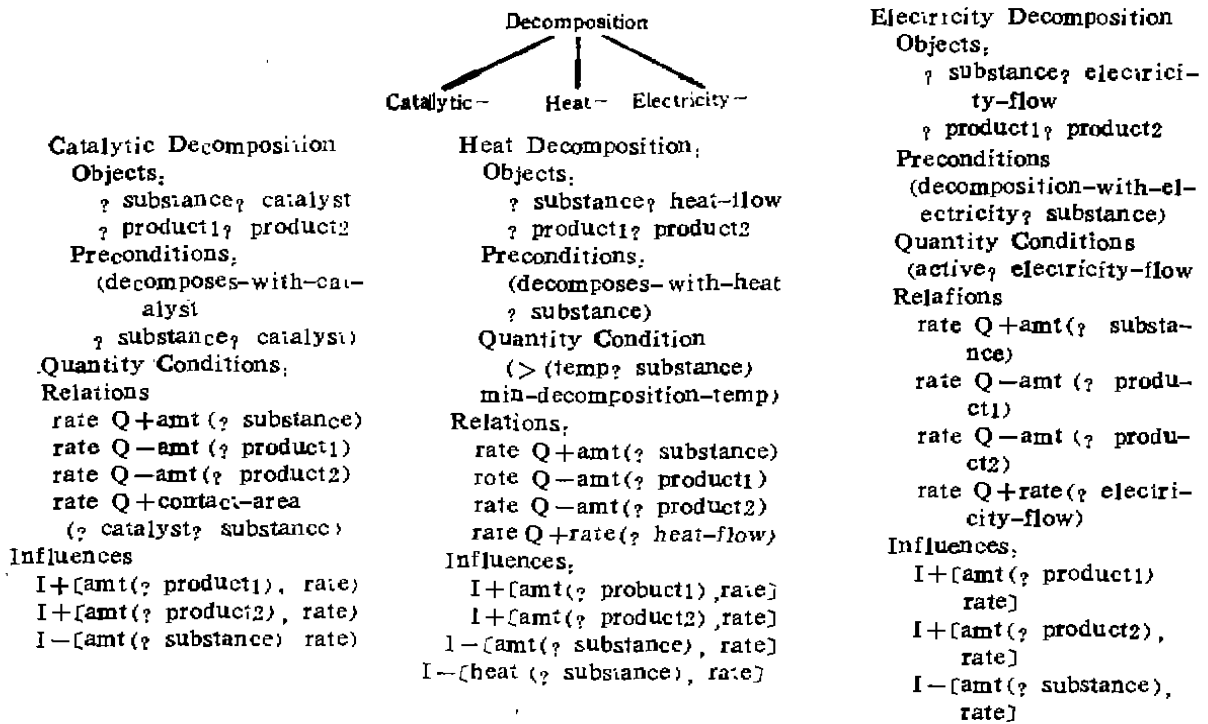


图3 化学分解的域论表示

在上面描述中，“A Q±B”记号表示A随B单调增加，单调减少，而“I±AB”记号表示A直接地，间接地受B影响。化学分解一般有三类：催化分解，加热分解和电解。这时我们可以建立化学分解过程的完整域论(图3)。

### 三、解释归约方法举例

多种解释问题一般起源于难处理的域论，在化学论域中存在许多这类问题，如：化学分解问题。前面我们已经建立了分解反应的定性理论，假设现在的任务是生产大量的氧气，然而初始的问题求解系统不具备产生氧气的方法，我们用图4表示从中提取氧气的一个脚本：

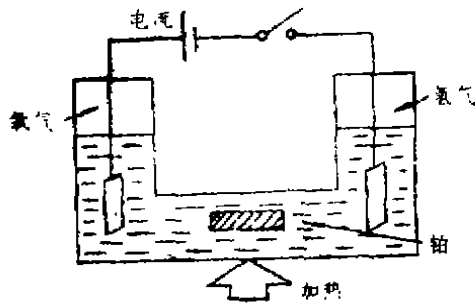


图4 产生氧气的脚本

首先,解释构造器试图使用一般化学分解理论来解释氧气产生的过程,它失败了,因为在通常的情况下,水不能分解成氢气和氧气。这时解释构造器搜索较为详细的分解理论,根据已知的化学分解类型,可以找到三种不同解释:

- 分解是因为受热
- 分解是由于通电
- 分解是由于加入铂作为催化剂

解释构造器把这三种解释作为输入,把它们输入到解释归约模块,通过试验推理机来排除二种假设。对于上面我们给出的试验脚本,求精和鉴别策略都无济于事,因为从推理机中所得到的推测不能排任何一种除假设,试验推理机这时转换脚本,把脚本转换成热量不断增加的新脚本。推理机推测在加热分解的条件下,对于新脚本氧气产生速度加快,因为分解速度依赖于供热数量,然而对于其它两种假设,反应速度将不受影响。我们在新、旧两种脚本下,把分解速度进行比较,当比较的结果是速度不变时,我们就排除了加热分解这种假设。

这时还剩下两种假设:催化分解和电解。我们再一次使用转换策略,把脚本转换成使铂与水的接触面积增大的新脚本,在这种转换下,催化分解假设将推测氧气在新脚本下产生速度加快,然而电解假设推测产生速度不变。我们在新、旧两种脚本下,观察氧气产生速度,当观察的结果是不变时,我们又推出了催化分解假设的可能性。

通过以上推理,试验推理机发现分解反

应起源于电流,这时,基于解释学习算法把该解释进行一般化推广,从而获得电解的学习描述。

#### 四、对解释归约方法的评价及其结论

试验设计是解释归约的中心议题,对试验设计系统的评价根据下面四个标准:

- 完全性 对于任何两种解释,试验设计系统都可设计一个试验来区分这两种解释。
- 效率 为了排除假设,试验设计系统所设计的试验数目达到最少。
- 容忍度 如果现有的试验不能获得任何信息的话,试验设计系统应当能够设计其它试验。
- 可行性 试验设计系统应当仅设计现实世界中可行的试验。

我们已经讨论了因不完善域论而产生的多种解释问题的解决方法,即解释归约。在基于解释学习过程中,当遇到该问题时,即可使用解释归约技术。解释归约是在机器学习和定性推理中使用的一般技术,除了难处理域论,它还可用于因不完全或不正确域论而产生的多种解释问题,同时,它也可用于定性推理中,由于定性推理的歧义特征,可能存在许多可能性,解释归约的目的就在于获取信息来排除与现实不符的可能性,进而使推理任务比较容易处理。

#### 参考文献

- [1] 陈火旺,张少平,“基于解释学习”,《计算机科学》,No.1,1990.
- [2] Shankor Rajamony & Gerald DeJong,“The Classification Detection and Handling of Imperfect Theory Problem”, IJCAI 87
- [3] A. M. Segre & G. F. DeJong,“Explanation Based Manipulator Learning: Acquisition of Planning Ability Through Observation”, Proc. of the IEEE International Conf. on Robotics and Automation, 1985.
- [4] M. Lebowitz,“Integrated Learning: Controlling Explanation”, Cognitive Science, 1986.