

基于解释的学习

陈火旺 张少平 (国防科技大学计算机科学系)

摘要

近几年来,在机器学习领域,人们对基于解释的学习方法产生了极大兴趣。本文首先通过一个例子说明该学习方法,然后作了一般描述,并把它与基于类似的学习进行了比较,最后详尽讨论了基于解释的学习中的两个重要方面:操作性和不完善域论。

一、引言

近几年来,在机器学习领域,又出现了一种新的学习方法,即基于解释的学习(Explanation-based learning)。最初,机器学习的研究重点集中在基于类似的学习(Similarity-based learning)上。基于类似的学习方法依赖于归纳推理,对一组实例进行一般性概括,从而产生一条实例分类规则。然而,该类方法不使用论域知识进行推理判断,来解释实例属于或不属于被学习概念的原因,而是仅仅通过实例之间的比较,以及简单的词法分析,归纳实例的分类规则。为了避免基于类似的学习方法这方面缺陷,基于解释的学习方法应运而生。基于解释的学习系统具有某一领域的知识,并且能够根据该知识,对实例进行详尽的分析,从而构造解释,产生规则。

二、基于解释学习的一般描述

首先,让我们从一个简单的例子着手,来说明基于解释的学习方法。假设我们要学习的概念是:一个物体能够被安全地放置在另一物体之上,而没有导致第二个物体的倒塌。我们已知的事实包括:一个实例,及一组公理和规则所构成的理论。实例说明了《人工智能原理》这本书能够安全地放置在书桌上,而理论则规范了放置物体的安全性。

然而,已知的关于放置物体的理论,不能作为识别一个物体放置到另一物体之上是否安全的有效描述。

假设根据给定的理论,我们能够证明出已知实例为我们要学习概念的一个例子,即《人工智能原理》这本书能够安全地放在书桌上,因为这本书要比书桌轻。我们把该证明进行一般化推广,即任何一本书都要比任何桌子轻。根据这一般性解释,我们就可以得出这样的结论:“任何一本书都能安全地放置在任何桌子之上”。这就是我们基于解释学习方法所要学习的最终描述。

虽然上面的例子很简单,但它却表达了基于解释的学习的基本思想;下面我们就基于解释的学习作一般性描述。

在基于解释的学习系统中,我们把要学习的概念称为目标概念(Target concept),它是概念的高级描述,目标概念描述了一类实例。对于基于解释的学习系统,系统的输入由二部分构成,一部分为域论(Domain theory),另一部分为训练实例(Training example)。训练实例是从目标概念中精选出的例子,如何选取训练实例,要根据不同的基于解释的学习系统而定,而域论则为描述某一论域的一阶理论,它由一组公理和规则组成。就现有的基于解释的学习系统而

言,学习都是针对某一特定的论域,且要求对于该论域,能够建立一个完善的域论,关于域论问题,我们还要在第四部分中详细讨论。基于解释的学习系统的关键是其学习算法。首先,学习算法使用域论知识,来解释为什么训练实例为目标概念的一个实例,解释的结果构成了训练实例满足目标概念的一个证明;然后,学习算法把训练实例的解释结构进行一般化,从而找到解释成立的最弱条件。我们把该条件称为目标概念的学习描述(Learned description)。基于解释的学习系统的最终输出就是产生目标概念的学习描述。学习描述既是训练实例的一般化,又是目标概念的特例化。学习的目的就是要产生学习描述,作为目标概念的有效识别器。

我们可以在逻辑框架下,对前面所举的例子,进行如下的形式化描述:

首先,我们的目标概念是:

Safe-To-Stack(Obj1, Obj2)

训练实例为描述物体Obj1, Obj2的下述事实:

On(Obj1, Obj2)
Isa(Obj1, book-AI)
Isa(Obj2, table-book)
Color(Obj1, red)
Color(Obj2, blue)
Volume(Obj1, 1)
Density(Obj1, 0.1)

而域论则为放置一个物体到另一物体之上的安全性规则:

Not(Fragile(y)) $\rightarrow$ Safe-To-Stack(x, y)
Lighter(x, y) $\rightarrow$ Safe-To-Stack(x, y)
Volume(p, v) $\wedge$ Density(p, d) $\wedge$ $\times(v, d, w)$
 $\rightarrow$ Weight(p, w)
Isa(p, table-book) $\rightarrow$ Weight(p, 5)
Weight(p₁, w₁) $\wedge$ Weight(p₂, w₂) $\wedge$ $\lt(w_1, w_2)$
 $\rightarrow$ Lighter(p₁, p₂)

基于解释的学习系统验证训练实例为目标概念Safe-To-Stack的正确例子,从而构造出如下图所示的实例满足目标的证明。

我们把该证明进行一般化,从而导致下

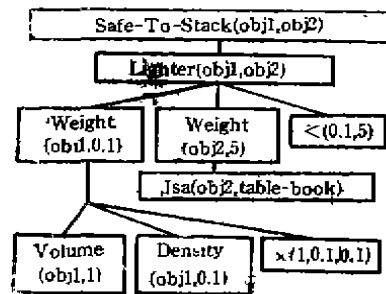


图1 Safe-To-Stack(Obj1, Obj2)的解释结构(证明树)

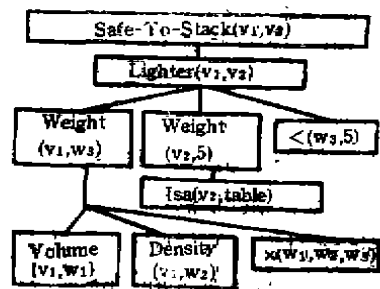


图2 Safe-To-Stack(v₁, v₂)一般化解释结构(证明树)

面Safe-To-Stack(v₁, v₂)的一般化解释结构。

一般化证明树叶节点谓词的合取,构成了目标概念的条件,从而产生一条归纳规则,即学习描述:

Volume(v₁, w₁) $\wedge$ Density(v₁, w₂) $\wedge$ $\times(w_1, w_2, w_3)$ $\wedge$ Isa(v₂, table) $\wedge$ $\lt(w_1, w_2)$
 $\rightarrow$ Safe-To-Stack(v₁, v₂)

三、与基于类似学习的比较

基于类似的学习是前一阶段人们较为重视的一种学习方法。该方法特别适用于分类任务,即判断例子是否属于某一特定的解类。在基于类似的学习系统中,学习算法通过对输入的训练实例集合中的实例进行比较,分析实例之间的共同以及不同特征,从而归纳出实例的分类规则。

然而在基于类似的学习中,算法的输入不是一个训练实例,而是由一组训练实例所组成的训练实例集合。为了能够使学习算法的输出,即归纳规则具有精确性,需要我们精选训练实例集合中的成员,同时,归纳规则

的精确性将随着训练实例集中的训练实例数目的增加而增加,但是,增加训练实例的数目将会产生一个不良的影响,即学习算法为了归纳出分类规则所需的时间花费将会显著增加,对于某些算法这种增长是指数级的。

其次,基于类似的学习要求训练实例必须是正确的,对于绝大多数的基于类似的学习系统而言,训练实例集合中不应含有不协调的实例。

我们知道,基于类似的学习系统不使用域论来分析、判断实例的属性,而仅仅通过实例之间的比较来提取共性与不同,这就不可避免地会导致所谓的归纳跳步问题,即对目标概念的不同属性的重要性不加以区分,而把它等同看待,因为系统不具有域知识来进行说明。让我们以恐怖主义事件的分类学习,来说明归纳跳步问题。

基于我们所掌握的一些恐怖主义事件,我们的学习任务就是要归纳出一条分类规则,从而根据该规则来识别恐怖主义者。假设在每一恐怖主义事件中,恐怖主义者都携带手枪且穿条蓝色牛仔裤。系统可能会产生一条这样的归纳规则:凡是携带手枪且穿有蓝色牛仔裤的人都是恐怖主义者。学习系统把携带手枪和穿牛仔裤这两个属性等同看待,而不加以区分。事实上,就恐怖主义事件而言,携带手枪才是最为实质性的特征,它要比穿牛仔裤这一特征更为重要。

通过以上分析,显然我们可以看出,基于解释的学习要比基于类似的学习更具优越性,我们可以把基于解释的学习的优点归纳如下:

- 基于解释的学习仅需要一个训练实例。
- 基于解释的学习并不要求训练实例为目标概念中实例的最佳选择,只要它能充分地说明目标概念。
- 如果域论是完善的,基于解释的学习系统所得到的学习描述就是正确的。
- 在基于解释的学习中,不存在归纳跳步问题。

然而我们需要指出的是,有些论域对于

基于解释的学习方法不太适合,例如医学诊断。因为对于这样的领域,我们很难建立一个完善的域论,用基于类似的学习方法解决这类问题可能较为恰当一些。

四、操作性的定义

在基于解释的学习中,为了产生目标概念的最终描述,学习系统必须具有其初始描述。事实上,如果不知道目标概念的初始描述,而要系统来解释为什么训练实例满足目标概念是不可能的。如果说目标概念的初始描述是基于解释的学习的先决条件,那么学习的意义何在呢?

问题的关键在于,学习包括知识转换(knowledge transformation),基于解释的学习不是获取“新”知识,而是转换非操作知识成为可操作形式知识。虽然给定的关于目标概念的初始描述是正确的,但是该描述是一种非操作(non-operational)形式的描述,而非操作描述意味着该描述不能被系统有效地使用。概念的非操作描述与操作描述之间存在着很大差异,基于解释的学习系统的任务就在于通过转换,缩短两者的差异。

虽然操作性是区分基于解释的学习中最终描述与初始描述的关键,但是许多学习系统并没有精确地定义操作性,并把操作性看成是独立的,静态的二值描述特征,事实上,操作性可能是动态的,多值的描述特征,它依赖于学习系统的执行任务及类型。

首先,我们来区分概念与它的描述。假设所有的实例构成了一个实例全域,概念则为实例全域的子集,概念描述为实例全域上的谓词,概念通过概念描述来详尽表示。如果两个概念描述指称同一概念,则我们称这两个描述是同义的,我们用下图来说明它们间的关系(见图3)。

图中的概念空间表示能够由学习系统所描述的所有概念,概念空间的每一点与实例空间的实例集合相对应。概念描述空间被划分成操作的和非操作的描述两个区域,概念空间与概念描述空间的点之间存在着多对多

概念描述空间 概念空间 实例空间

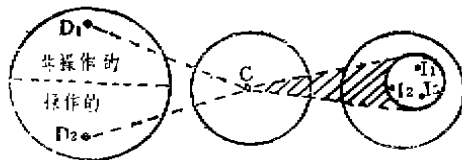


图3 概念和概念描述间的关系

的对应关系。例如， D_1 和 D_2 是两个同义描述，两者皆描述概念 C ，概念 C 包含 I_1 、 I_2 及 I_3 等实例， D_2 为 C 的操作的描述，然而 D_1 为非操作的描述，因此当描述同一概念存在不同的方法时，操作性定义了一种描述比另一种描述优越的准则。

操作性在基于解释的学习中的作用，可以通过把基于解释的学习看成是在概念描述空间中进行搜索来加以理解。例如，在图中 D_1 表示提供给学习系统的初始的非操作的描述，而 D_2 则为最终学习到的操作描述。我们可以把 D_1 看作搜索的初始节点， D_2 看作解节点，解释作为在描述空间的搜索方式，而操作性则作为搜索终止的标准。

现在我们来定义操作性。如果我们称一个概念描述为操作的，它需要满足下面二点要求：

- 能用性 (useability)，它能够被学习系统用来识别它描述的概念实例。
- 效用性 (utility)，当学习系统使用该描述时，系统的执行性能应该得到改善。

那么操作性又是如何确定的呢？实际上，每个基于解释的学习系统都含有一个过程来确定操作性，该过程评价概念描述，从而产生一个操作性度量作为它的输出。

在基于解释的学习中，操作性具有下述三个特性：

- 变化性 (Variability)，它是指操作性将随着时间的变化而变化。随着学习的不断深入，执行环境的变化，初始为非操作的描述可能会变成操作的描述，反之亦然。

- 粒度 (Granularity)，它是指操作性的度量。如果过程的输出为操作的和非操作的两种，那么度量称为二值度量，否则就称为连续度量。连续度量的优点在于，学习系统可以估计操作性的程度。

- 确定性 (Certainty)，它是指操作性估价过程所产生的操作性度量的可信度。很多系统不能保证它所产生的估量的精确性，换句话说，系统不能实际测试描述是否对识别实例有效。

综上所述，操作性在基于解释学习中，既作为区分初始概念描述和最终概念描述的关键特征，同时也作为基于解释学习系统的学习准则。

五、不完善域论问题

正如我们前面所指出的那样，如果域论是完善的，基于解释的学习系统所产生的学习描述就是正确的。然而在实际中，建立的域论为不完善的事情在所难免。这就需要我们的学习系统具有自动地测定和改正不完善域论的能力。下面我们就不完善域论进行简单分类，并且讨论一下在基于解释的学习系统中，不完善域论对解释的影响。

不完善域论大致可以分成以下三种：

- 不完整的理论 (incomplete theory) 由于缺少一些公理或规则，某些演绎不能完成。
- 不协调的理论 (inconsistent theory) 不协调的结论可以从该理论中导出。
- 难处理的理论 (intractable theory) 演绎由于受到计算的复杂性限制，而不能完成。

然而有时很难把三者严格地区分开来，例如，有的域论可能既是不协调的理论，同时又是难处理的理论。

在基于解释的学习中，构造解释是指使用域论的事实和规则，来说明为什么训练实例为目标概念的实例，我们可以用下图来说明解释构造过程：

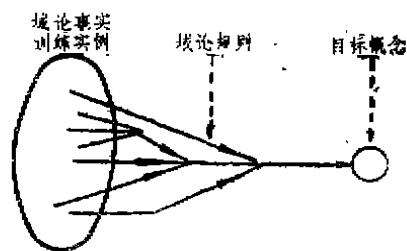


图4 解释构造过程

使用域论中事实和训练实例，使用域论规则，一步步地演绎，最终构造出训练实

例满足目标概念的证明。然而由于不完备的域论,在解释构造过程中,可能会遇到下述问题:

- 解释中断(Broken Explanation):在域论中,由于缺少必要的事实和规则,解释构造被迫中断,如图5(a)所示。

- 矛盾(Contradiction):系统在构造解释过程中产生了矛盾。这类问题可能由域论中包含了错误的事实和规则,如图5(b)所示。

- 多种解释(Multiple Explanation):当对某一结论仅期望一种解释的情况下,系统为该结论构造了多种解释,这类问题是由于缺少知识来帮助区分不同的解释,多种解释如图5(c)所示。

- 资源超界(Resources Exceeded):在构造解释过程中,系统所需资源(时间、空间)超越了分配给它的界限,这类问题可进一步划分成:

- 大空间搜索问题(Large Search Space Problem) 系统在构造解释过程中,必须搜索较大的空间,虽然解释存在,但系统却构造不出来,因为存在太多的选择路径,系统不具备知识来进行路径

的选择,而不得不搜索所有的路径,如图5(d)所示。

- 小连接问题(Small Links Problem):在构造解释过程中,构造的步幅太小,步数太多,以致在分配的资源之内,无法完成解释(难处理问题),该类问题与大空间搜索问题无关,可能在不涉及搜索的情况下该类问题同样存在,如图5(e)所示。

为了处理上述问题,需要获取新的知识,一种叫做试验设计方法可以应用于上述各类问题的处理,我们在这里就不详尽地讨论了。

六、结束语

基于解释的学习方法作为机器学习的一种新方法,自创立以来,已经有了很大的进展。现在已可应用于较为复杂的领域学习,如数学、电路设计、自然语言、机器人、物理及化学等十分广泛的应用领域。由于基于解释的学习的基础机制已经很清楚,现在人们把注意力转向了独立于特定论域的基于解释的学习算法的研究。我们相信,基于解释的学习方法一定会有光明的应用前景。

参考文献

- [1] Haym Hirsh, "Explanation-based Generalization in a Logic Programming Environment" IJCAI 87
- [2] Shankar Rajamoney & Gerald DeJong, "The Classification, Detection and Handling of Imperfect Theory Problems" IJCAI 87
- [3] J. L. de Siqueira N. & J. -F. Puget, "Explanation-Based Generalization of Failures" ECAI 88
- [4] D. Dubois & H. Prade, "Defining Operationality for Explanation-Based Learning" AI, vol. 35, No.2

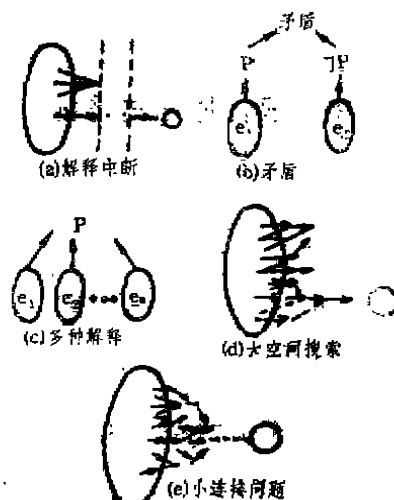


图5 解释构造中的问题