

基于事例的推理导论

Ralph Barletta

Barletta R 田伟涛

TP18

当选择医生看病时，我们更愿意选择年长的，为什么呢？固然，刚毕业的医生有最新的诊断和治疗知识，而且比有经验的老医生有更多的前沿书本知识，但我们更愿意选择老医生，因为老医生已经见过和治疗过许多与我们有类似疾病的病人。从本质上说，我们考虑医生的经验时，更多的是根据他们曾处理过多少病例，而不是他们懂得的治疗知识。基于事例的推理(Case-Based Reasoning, CBR)是人工智能技术的最新进展，符合专家迅速、准确地求解“新”问题的过程，因而这种从过去的相关事例推理的思想非常吸引人。

从选择医生可以看出事例经验在指导专家解决问题中的作用，准确地检索和操纵过去求解问题的事例的能力在许多领域(诊断、

分类、预测、规划、设计、法律、过程控制与监测、高级制造业配置等)中是很重要的。

计算机不会健忘，如果让它们记住正确的东西，我们就可以享受到我们自己的或他人的过去成功或失败的经验，这就是成长中的CBR研究界的最终目标。

一、CBR方法

基于规则的专家系统解决问题时，采用一个输入说明(或者通过与用户的问答开发出一个输入说明)，然后从规则库中将适宜的规则集联结起来，从而得到问题的解答(图1)。若给出完全相同的问题情况，这个系统将进行完全等量的工作来得到解答，也就是说，基于规则的系统不会学习。另外，若给出问题定义域之外的问题(例如同领域内

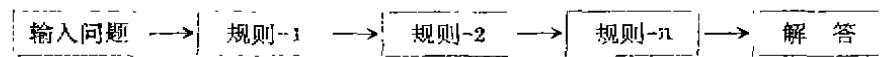


图1 基于规则的专家系统求解问题的过程

新的问题类型)，系统通常不能得到任何解答。最后，基于规则系统的建立和维护是非常耗时的，因为从专家身上抽取知识是一种劳力密集型活动，而且规则之间互相依赖，使系统添加新知识时调试任务十分复杂。

CBR系统以一种完全不同的方式运行(图2)。若给一个输入说明，CBR系统将搜

索它的事例库以寻找一个匹配输入说明的现有事例，如果幸运的话(幸运将随着向系统添加新事例而增加)，将找到完全匹配输入问题的事例而直接得到问题的解答，这就使迅速解决复杂问题成为可能。另一方面说，不那么幸运的话，也将检索到一个与输入说明相似的事例，但不能作为完全适宜的解答。

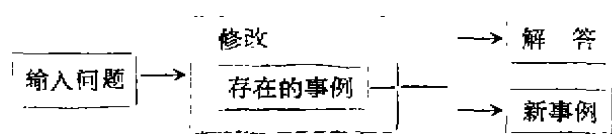


图2 CBR系统搜索它的事例库

CBR系统(或用户)必须发现并修改所检案的事例与输入说明不相匹配的那部分,这叫做事例改写。事例改写的结果是一个完全的解答,同时生成一个新事例并自动地加入系统的事例库以备将来使用。(学习是CBR系统的一个基本部分。)这样,若将来给出相同的问题,系统将越过它第一次解决问题时所做的工作。总之, CBR系统容易建立和维护,因为它把知识工程的任务减化为定义术语和从专家那里收集事前分过类的事例,并把增添新知识简化为向事例库增加一个新事例。

二、CBR的起源

大多数AI工作者认为, Roger Schank在1982年出版的《Dynamic Memory(动态记忆)》著作中,描述了CBR的最早研究工作,给出了计算机上建造这种推理系统的方法。他的早期思想被他的一些耶鲁大学学生所推广,1983年,很著名的佐治亚工学院的Janet Kolodner的Cyrus系统在计算机上首次实现了Schank著作中的许多原理。尔后,耶鲁大学、佐治亚工学院和马萨诸塞大学的研究生们又建造了一些系统,这些系统在法律、烹调、医药等领域中证明了CBR。

CBR先驱们的工作确定了基本的研究领域和问题,这些问题正在被研究。随着CBR研究的成熟, CBR正投入工业应用。

自1987年以来, CBR研究界每年都举行了研讨会,召集这一领域的研究者,参加者对CBR的兴趣大大增加,其中三个研讨会出版了会议论文集,对CBR有兴趣的人来说是极好的读物。

三、CBR研究课题

大多数CBR研究分如下几大类:事例表示,索引,存贮和检索,改写,学习和归纳。

3.1 事例表示

什么是一个事例?怎样向计算机描述一个事例?就最简单的形式来说,事例就是能导致特定结果的一系列特征,例如,个人贷款表以及贷款是否被批准,病历及其相应诊

断。就最复杂的形式来说,事例就是形成问题求解结构的子事例的关联集合,例如,一架飞机和一个电路的设计由许多子设计组成,组成整个设计问题的每一部分对整个问题的来说可以认为是一个事例。

CBR系统中主要的知识工程任务是确定适当的事例特征,包括领域专有名词的定义和专家借以解决问题的代表性事例的收集。基于规则和基于事例的知识工程之间的区别在于,自动的事例索引技术大大减少了从专家提取和构造规则型知识的必要,而这正是基于规则的知识工程中最费时间的部分。

3.2 事例索引

CBR系统的强大功能来源于它能从其记忆库中迅速准确地检索出相关的事例。事例索引过程的目标是弄清在将来相似的情形下,在什么时候应该检索一个事例。事例记忆库和检索过程的目的是建造一个结构或过程来得到最适当的事例。事例索引技术通常有三种:最近邻法、归纳法、知识导引法,或者这三种方法的结合。

最近邻法是指用户利用与记忆库事例相匹配的输入事例的特征权数和来检索事例。最简单的情况是所有的特征具有相同的权值,系统将选择匹配八个特征的事例而不是匹配六个的事例。如果检索目标未能很好定义或可用事例不多,这一方法是不错的。单独使用这一方法的最大问题是不可能覆盖全局特征的权值集,在所有情况下难以准确检索到事例。许多问题的特征权重是相互依赖的,在确定要检索的适当事例时,一个给定特征的重要性取决于这个事例的其它特征的权值。

如果检索的目标或事例的结果是良定义的,并且有足够的事例进行归纳比较,那么归纳索引方法比最近邻法好。前者的任务是在需要被分类的各种各样的事例(例如好贷款与坏贷款)之中,归纳地确定哪一特征判别能力最好。

归纳法有两个优点。第一,能自动、客

观、严格地分析事例，确定能区别这些事例的最佳特征。第二，事例可以组织成分层结构供检索用，其检索时间成对数而不是线性增加。当使用的事例库包括数以千计的事例时，检索时间就成为一个重要的因素。然而，为了完成归纳，系统需要相当数量的事例来生成判别特征。归纳分析时间长，是归纳索引法的主要缺点。

基于知识的索引法尝试利用现存的有关事例库事例的知识来确定检索事例时哪些特征是重要的。如果这样的解释性知识是可用的并且可表示的话，这种方法是可取的。问题是，常常难以代码化足够的解释性知识，在大范围的可能事例输入上完成完备的基于知识索引。因此，许多系统使用与其它索引技术相结合的基于知识的索引法。对多数现实领域来说，索引知识是可用的，重要的是CBR系统要去利用。

3.3 事例存贮和检索

一旦事例表示好了而且索引好了，就可以被组织成一个高效的检索结构，大多数事例存贮结构介于纯粹的关联检索与纯粹的分层检索之间。前者指一个事例的一些或全部的特征与其它的特征相独立地索引；后者指事例特征高度地组织成一般到特殊的概念结构。最近邻匹配技术被认为是关联的，因为它们没有真正的记忆组织；判别网络(Discrimination net)主要介于关联与分层之间，具有网结构的检索灵活性，因为它们有许多潜在的索引特征链；决策树是一种纯粹的分层记忆组织。

存储组织的类型关系到执行索引可用的知识量和系统的检索要求。要求灵活性时，采用关联的方法；检索任务很好定义时，采用分层的方法，因为可节省更多的检索时间。

3.4 事例改写

事例检索的目标是得到一个与输入情况最相似的过去事例。事例改写是指将一个检索到的与当前情况大部分匹配的事例改写成

完全匹配的事例，事例改写也包括做最小的改变，使输入要求匹配一个已知的所存事例。确定一种普遍可用的改写方法是困难的，因为改写要面向特定的问题，一种问题（如设计）的改写与另一种问题（如诊断）的改写是不同的。

大多数现有的CBR系统为特定的问题领域实现改写的方法是将改写知识代码化为改写规则集或领域模型。改写规则可使检索到的事例转化为一个满足所有要求的新事例。许多最近的应用已成功地利用存贮的现有事例来完成改写，在很难将改写知识表示成规则形式的问题领域中，利用事例是最好的（纵然不是唯一的选择）。而且，即使事例不能被计算机改写，至少这个系统已经为人类“改写者”提供了一个有意义的起点。

3.5 学习和归纳

学习和归纳对CBR系统来说是重要的，利用现有技术从事例中抽取有用信息，使CBR方法避免了基于规则方法的主要问题，即收集问题求解或分类的知识并将它们正确利用。开发中的基于归纳和解释的索引技术将使CBR系统得出有用的索引特征，并将它们构造于高效的存贮组织之中。随着事例的积累，事例归纳可以确定出体现特定事例群主要特征的原型事例，并且原型事例可与特定事例存贮在一起，提高系统长期运行的准确性。

四、CBR与其它AI技术的比较

将CBR与其它知名的和正出现的以下技术相比较：基于规则的系统、神经网络和模式识别技术等。

4.1 CBR与分类任务

在分类问题上CBR比基于规则的方法有一些优点，它将节省从专家那里抽取完备、一致的分类规则的时间，从要分辨的每一类中收集代表性事例，并分类生成可描述特征的适当集合。这一工作也是基于规则系统必不可少的一部分，因而开发CBR系统在这方面并不增加显著的时间。

CBR系统将节省把所得规则结构化到功能系统的时间,即使有专家系统外壳工具用于知识结构化,这仍然很困难和费时,尤其是在复杂的分类领域。在某些CBR外壳里,归纳学习和索引过程可以自动地建立分类的最优决策树。

把规则嵌入到现有系统中去是很困难的,这将影响维护和改进这个系统的长期费用。而向一个CBR系统中添加知识时,只需添加更多的事例并让归纳索引机制重新将分类树结构化。若需要改变CBR系统的分类目录时,只需将现有分类改为另一种分类,再运行索引机制即可。但在基于规则系统中,开发者就必须从头重做这一系统。

在解释和证明所得分类方面, CBR也优于基于规则的方法。专家系统中的规则是事例中信息的提炼,基于规则的系统仅仅显示给用户分类的一系列规则,最初的事例并没有用。而CBR系统证明它的分类不仅显示决定分类的规则(通过索引过程),而且显示支持所用索引的事例。在不能决定分类时,让用户看一看真实的事例将有助于做出最终的分类。在既不能用计算机索引事例,也不能用计算机激活规则的情况下(如图象、声音和自由格式文本), CBR系统仍可使用户存贮有用的分类信息。

4.2 CBR与神经网络和模式识别

模式识别技术用于分类问题已好多年了,其基本方法是采用“目标概念”事例和相反事例的集合(如好或坏的贷款),并产生一个数学函数来确定未见过的输入类。现有许多种模式识别算法,最流行的是线性或二次判别和贝叶斯(概率)分类。

线性或二次判别方法的最终输出是数学函数,因而它仅在输入事例特征是用数字描述的情况下才能正常工作,在具有重要非数字特征的领域中,这大大限制了模式识别方法的用途。贝叶斯分类法的问题则很明显,它需要对样本中相互依赖的变量给出完全的概率估计,对大多数现实领域而言,这很难

做到。

即使你建立了一个准确的判别函数,但对新事例只能做出一个或黑或白的结论。没有事例的支持证据,这就限制了用户与过去事例进行比较,做出更深的结论。模式识别方法的不足已成为这些年来发展符号机器学习 and 神经网络方法的主要动因。

神经网络方法以其能力赢得了注意。它的准确的分类结果与符号归纳方法相比时好时坏,它擅长于处理“噪音”数据——含有错误的的数据。神经网络一旦训练好,就会快速地分类,这使它很适于实时应用。

除了优点之外,神经网络也有许多缺点。它应用于广泛的分类问题时很困难或不现实。有代表性的是,用布尔或数字向量来定义事例的特征是很困难或不可能的,因为复杂事例之间具有内在的关系结构,而不是简单的罗列。计算方面,神经网络需要大量的数据和CPU时间来得到一个准确的网络结构,对于复杂的数据集,这意味着要花几周的时间来训练网络。

神经网络的结构对其分类能力有巨大影响。网络建造者在决定网络结构时非常自由,可以改变节点数、节点间的联结、网络节点的层数,但这些灵活性带来了问题,现在没有针对所给问题领域需要而设计神经网络的确切方法,多半是一种艺术而不是一种科学。系统开发者不得不尝试不同的网络结构来寻找最佳者。训练需要大量计算时间,而且采用试错法,这增加了用神经网络方法开发分类系统的困难。

即使表达好事例,训练好网络,建立了满意的网络结构,但给出一个新事例时你也不明白这个分类是如何得到的,解释的方法还没有,你只能知道节点、联结和权重。

CBR系统的归纳索引能力是它优于神经网络和模式识别技术之处。与后两种技术相比,归纳系统可以表达和学习范围更加广泛的特征类型。CBR使用大量事例特征集的能力,其准确性优于其它方法或至少一样,而

所需事例的数量和训练归纳系统的时间少于模式识别，并大大少于神经网络。归纳系统的输出具有较高的可读性，这就减少了额外的解释培训。最后，CBR系统可以完成更多种类的任务，不仅仅是分类问题，而神经网络和模式识别则不能。

4.3 CBR与复杂合成问题的基于规则系统

基于规则系统用于解决复杂的现实问题，如规划、调度和设计，它的成功原因在于三方面；专家了解该应用领域，领域行为可表达，问题是易处理的。

基于规则方法解决规划或调度问题的困难在于必须“从头开始”，对于现实问题这很消耗计算资源，这个问题并不是因为专家系统中的规则激发不太高效，而是因为从头开始建造复杂的规划要涉及大量的搜索（规则激发），系统需要面临数以千计的选择。

“跳过开始”的方法很必要。CBR系统解决象规划、调度和设计等复杂问题时要靠寻找一个类似的、成功的过去事例，并修改它以符合当前问题的所有需要。这要求系统拥有相当好的事例集以供使用，假如不能从过去的事例中得出一些有价值的东西，CBR还有何用。然而，没有多少领域是这种情况，即使是强调革新的设计领域，工作也是从过去的成功做起的。

由于大多数系统检索到的事例不符合当

前情况的严格要求，因而自动的事例改写技术是CBR中的关键问题。大多数CBR方法依靠改写规则来进行改写，这就使CBR与基于规则系统的结合成为可能。

CBR方法擅长于那些并不需要去很好理解的领域，因为系统不需要知道一些事情过去是如何做的，它只需知道在什么情况下这一方案将来可以重新使用。CBR系统擅长于那些表达困难的领域，而在这些领域内基于规则的方法是行不通的。事实上，CBR系统可以展示出以前并不知道的领域特征间的关系，它所拥有的归纳技术使“理解”问题变得容易。

许多复杂问题并不绝对地归于基于事例或基于规则的系统，如经济或政治分析、建造太空飞船等问题，但CBR至少可以帮助规划者利用过去的经验。

五、一条新道路

一位干过20年的医生也许并不知道刚毕业医生所知道的全部技巧，但是当你不知道痛从哪儿来时，医生的经验是最重要的。就象我们选择一位有经验的医生一样，CBR是一项非常吸引人的技术，因为它符合人类专家解决问题的过程。在现实中，CBR不仅仅是可能的，而且通常是更好、更快和更准确地解决问题的方法。（参考文献略）

〔田伟涛摘译自《AI EXPERT》，Aug. 1991，pp43-49，田盛丰校〕

*****下期内容预告*****

（常识的表示及推理专辑）

缺省推理与认识进程；容错推理；认知硬度与默认分层理论；关于类比推的理若干基本问题的研究；模糊逻辑程序设计基础；知识/常识的表示和推理；智能的限定说；Bacon-Mill 实验推理方法的一个推扩；一种决策论和人工智能相结合的非确定性规划方法；NDBS：一个非单调数据库系统。