

57-60, 26

人工智能

决策论

非确定性

规划

10

计算机科学1993 Vol. 20 No. 2

一种决策论和人工智能相结合的 非确定性规划方法

TP18

祝明发

(国家智能计算机研究开发中心)

(中国科学院计算技术研究所, 北京100080)

摘 要

This paper proposes an approach to planning under uncertainty by combining artificial intelligence and decision theoretic methods. Goals are defined in decision theoretic framework. Actions are selected by Maximum Subjective Expected Utility rule. A causal model is established to compute probability distribution over world states. A planning procedure is given.

一、前言

规划是人工智能中的一个重要分支。一个规划是一个动作序列, 其执行的结果将世界从初始状态带到目标状态。在规划研究的历史中, 一个重大的突破是七十年代初的STRIPS, 它在解决“画面”问题(frame problem)中进行了成功的尝试。在STRIPS中, 一个动作由一组条件、一个增加表和一个删除表描述。动作的后果包含在这两个表中。任何命题或谓词, 只要不在这两个表中, 则不受该动作的影响。

这样, 就不需要显式画面公理来规定那些不受动作影响的条件。

STRIPS及其他类似的规划系统的一个固有的限制是, 它们基于这样的假设, 即, 规划机构具有世界所有有关方面的完全知识, 并能完全预言其动作的后果。因此, STRIPS式的规划是一种确定性规划。这种规划是理想的而不是现实的, 因为在多数情况下, 机构很难获得关于世界的完全知识和完全预言动作的后果。确定性或非确定性不是环境固

因此, 消除矛盾后的P的一致解可由式(3)求得, 将所有关于P的一致解环境组成的集合定义为PRMS下的标号概念, 并可证明基于该标号的P的解是一致的、完全的, 该标号仍具有ATMS中标号的四个性质, 详见文[4]。

基于上述思想, 可得PRMS原理:

• Procedure PRMS

- (1) if NOG或KB变化, 转(3);
- (2) 否则, 结束。
- (3) 对系统nogood分类, 确定每个谓词P的矛盾公式及极小nogood集NOG;
- (4) for 每个P do {

(5) 计算P的标号;

(6) 基于标号求解;

(7) 结束

6. 结束语

将非单调推理引入数据库是近年来人工智能与数据库相结合研究的前沿课题, 已引起国际上许多学者关注。我们以Reiter的缺省逻辑作为知识表示及其处理的基础, 针对当前的信念维护仅能解决命题级维护问题, 提出了一种面向谓词的信念维护系统PRMS, 有效地扩充了ATMS中标号的概念, 解决了实际系统中大型数据库、知识库系统中的信念维护问题。进而设计并实现了非单调数据库系统NDBS, 目前该系统已建立微机上, 用Quick C及Oracle语言共同实现。(参考文献略)

祝明发 获美国韦恩州立大学博士学位, 研究员, 主要研究并行处理和人工智能。

有的特性,而是机构的与环境有关的知识状态的特性。

在非确定性规划中,机构关于世界知识状态的信念可以用世界状态概率分布来表示。决策论框架中的规划方法在优化动作选择方面很有效,但不能简单地用来达成目标,因为它没有目标的概念,而且世界状态和动作之间不允许有因果关系。一些学者对信念网(或叫 Bayes网)进行过研究。一个信念网是一个非循环图(DAG),其结点代表随机变量,结点间的连线表示概率关系。为了简化因果关系推理,在DAG中找出条件独立性关系。基于这个思想,M.P.Weilman研究了STRIPS假定,并得出结论:机构关于那些没有明白描述的动作后果的信念(有条件地)独立于动作本身,而只依赖于那些显式描述的后果和背景信息。这些工作对因果关系的认识和因果推理的实现有一定的帮助。

二、动作选择的优化

决策论的一个主要内容是相对于可能性和机构的意愿如何优化动作的选择。因此,规划可以看作是一个顺序决策过程。可能性的程度可以用来表示机构的知识状态,而这种知识状态又是可以不断更新的以便反映动态的环境。动作选择的最优化是根据极大地满足机构意愿的可能性来定义的。决策者关于他的行动的可能输出的意愿是用效用来表示的。效用是赋给动作输出的标量。决策论用概率来表示可能性或部分信念,而信念是规划机构的认识状态。这种概率的主观解释对表示机构关于环境的知识状态是适当的。

Savage关于非确定情形下的选择理论是迄今为止最好的一种主观Bayes决策论。这个理论包括如下要素:

- (1) 作为决策机构所涉及的客体世界;
- (2) 一个世界状态集合 S ,每个世界状态是世界的描述;
- (3) 一个动作后果集合 E ;
- (4) 一个动作集合 A 。一个动作是一个从 S 到 E 的函数,即, $e=a(s)$, $e \in E$ 且 $s \in S$;
- (5) 将效用赋给动作后果的效用函数;
- (6) 一个二元关系 $\preceq$,它标明机构对

动作的偏好。 $a_i \preceq a_j$ 表示在两个动作中机构偏好动作 a_j 或者无偏好。

Savage选择理论的基本思想是,协调的世界状态概率分布和对动作后果的协调的效用分配,使得选择那些使期望效用极大化的动作正好反映机构的偏好。

令 $S=\{s_1, \dots, s_n\}$, $A=\{a_1, \dots, a_m\}$, 且 μ_{ij} 为状态 s_j 下 a_i 后果的效用,即 $\mu_{ij}=\mu(a_i(s_j))$, $i=1, \dots, m$, $j=1, \dots, n$ 。

主体期望效用定义为:

$$SEU(a_i) = \sum_{j=1}^n p(s_j) \mu_{ij}, \quad (i=1, \dots, m) \quad (1)$$

此处 $p(s_j)$ 为状态概率分布。

Savage的理论包含若干公理,由它们可推出:

(1) 存在一个唯一的概率测度 p ,使得状态 s_i 被认为比状态 s_j 更可能出现当且仅当 $p(s_i) \geq p(s_j)$;

(2) 对上面给定的 p ,存在一个从 E 到 R (实数域)的有界函数 μ ,使得 $a_i \preceq a_j$ 当且仅当 $SEU(a_i) \leq SEU(a_j)$ 。

如果一个机构选择一个使其主体期望效用 $SEU(a_i)$ 极大化的动作,则称该机构动作理性化。

对动作、状态和动作后果有两个重要的限制。第一,所有与机构做决策有关的因素必须由动作后果显式规定。第二,世界状态必须在因果关系方面独立于动作,即二者无因果关系。就决策而言,机构没有必要知道所有后果,所以,第一条限制不会产生大的问题。做决策所必需的后果只是集合 E 的一个子集,需要显式地规定。第二条限制意味着机构只能利用环境而不能影响环境。这与关于状态和动作间存在因果关系的人工智能理论相矛盾。如果动作的执行不能影响环境,则一个规划的执行不能使世界由初始状态过渡到目标状态。本文第四节将要详细讨论与第二条限制有关的问题。

三、构造效用函数

可供选择的动作的最优选择是依赖于效

函数的。在规划中,效用函数的构造必须反映机构达成目标的愿望。为了达成目标,与机构决策有关的所有动作后果必须显式地由动作的说明表(增加表与删除表)所规定。在非不确定性情形中,这种规定并不意味着机构确定地知道动作执行后的新状态。机构仅有关于世界状态的一定程度的信念(概率)。

设动作 a_i 的后果集合为 E_i ,且 $E_i'=\{e_{k_1}, \dots, e_{k_t}\}$,需要说明的所有动作的全部后果的集合为

$$E=\bigcup_{i=1}^m E_i'=\{e_1, \dots, e_t\} \quad (2)$$

动作后果集合可形式地定义如下:

$$E=E_1 \times E_2 \times \dots \times E_t=\{(e_1, e_2, \dots, e_t), e_i \in \{0, 1\}, i=1, \dots, t\} \quad (3)$$

注意, $k_i \leq t$, 即一个动作只能影响 t 个变量中的一部分变量。效用函数是一个在 t 维平行多面体域中无量纲标量函数。即

$$\mu=f(e_1, \dots, e_t) \quad (4)$$

此处 μ 表示动作 a_i 在状态 s_j 下执行的后果的效用,为简单计,下标 i 和 j 暂时略去。效用函数构造或赋值应满足这样的要求:规划机构在动作 a_i 和 a_j 间偏好 a_i 当且仅当对应于 a_i 的效用函数值大于对应于 a_j 的效用函数值。这种偏好顺序记作 $<^j$ (严格偏好)或 $\leq^j$ (偏好或无偏好)。在特定状态下找到一个最佳动作等价于找到效用函数的最大值。

偏好独立性概念在决策论中有着重要作用。如果在其他变量的值相对固定不变的情况下改变一个变量的值引起偏好加强,则称这个偏好结构独立于其他变量。例如,一个相对于 E_1 的偏好次序独立于 $E_2 \times \dots \times E_t$,当对于一定的 $\bar{e}_2, \dots, \bar{e}_t$ 下式成立:

$$f(e_1', \bar{e}_2, \dots, \bar{e}_t) < f(e_1'', \bar{e}_2, \dots, \bar{e}_t) \quad (5)$$

总蕴涵:对于任意的 $e_1, \dots, e_t$,有

$$f(e_1', e_2, \dots, e_t) < f(e_1'', e_2, \dots, e_t) \quad (6)$$

相对其他 E_i ($i=2, \dots, t$), 偏好独立性可以同样定义。

考虑两个动作 a' 和 a^0 , 它们的后果在 $E_{i_1} \times \dots \times E_{i_k}$ 上, 即, 它们只影响这几个变量。

现令 $(e'_{i_1}, \dots, e'_{i_k})$ 为 a' 的后果, $(e^0_{i_1}, \dots, e^0_{i_k})$ 为 a^0 的后果。 a' 和 a^0 可能有些相同的后果,但不会完全相同。若

$$f(e'_{i_1}, \dots, e'_{i_k}, e_{i_{k+1}}, \dots, e_{i_t}) < f(e^0_{i_1}, \dots, e^0_{i_k}, e_{i_{k+1}}, \dots, e_{i_t}) \quad (7)$$

对任意(在两函数中固定的) $e_{i_{k+1}}, \dots, e_{i_t}$ 成立, 则称相对于 $E_{i_1} \times \dots \times E_{i_k}$, 偏好次序 $(a' \leq a^0)$ 独立于 $E_{i_{k+1}} \times \dots \times E_{i_t}$, 并称这种

偏好独立性为成组偏好独立性。注意, 为了叙述方便, 效用函数中变元的次序做了调整, 即 a' 和 a^0 所影响的那些变元排在一起了。成组偏好独立性只与显式说明了的 a^0 和 a' 的后果有关。这并不是一个严重的限制, 因为规划机构对那些没有明确说明的不影响偏好次序的后果不感兴趣。任何与机构决策有关的后果必须显式地说明。

偏好独立性概念有助于简化效用函数的构造。T.C.Koopmans证明, 基于相当一般的假定, 相对于效用函数的所有变元对的顺序成对偏好独立性可推出相关的效用函数的可分离性或可加性。A.S.Tanguiane等讨论了效用函数的近似构造问题。就规划而言, 效用函数应该这样构成以使得较好的候选动作将世界带到离目标状态更近的状态。一个目标状态是一个包含目标为其世界描述的一个子集的状态。由此, 我们可以使动作 a_i 在状态 s_j 下的后果的效用与这些后果所满足的子目标的数目成正比。这种子目标数目记为 Q_{ij} 。若现行状态 s_j 离目标状态越近, 可能的新状态也应离目标状态越近。 s_j 与目标状态之间的“距离”, D_j , 定义为没有被 s_j 满足的子目标数。因此, 我们可以将效用函数定义为

$$\mu_{ij} = \beta(Q_{ij}/D_j) \quad (8)$$

注意这里 D_j 只与 s_j 有关而与 a_i 无关。若 $D_j=0$, 目标已达成, μ_{ij} 可赋予任何大于等于0的数。这时, 不需要进一步规划。参数 β 与动作 a_i 的条件有关。若 a_i 的条件全满足, 则置 β 为1。

反之,置 β 为0。主体期望效用可写作

$$SEU(a_i) = \sum_{j=1}^n p(s_j) \beta(Q_{ij}/D_j) \quad (9)$$

此处 $i=1, \dots, m$ 。上面讨论的效用函数是相对于动作 a_i 和一个特定的状态的。而规划机构并不确切地知道这个状态而只知道世界状态概率分布。因此,主体期望效用是机构意愿的一个恰当表示。

四、目标、因果关系和规划步骤

在传统的人工智能规划和决策论规划间存在着根本性冲突:前者以目标为中心概念,机构通过规划的制定和执行来达成目标;而后者没有目标的概念,最好的规划是使主体期望效用达到最大值的规划。在决策论中,机构的意愿由对规划中动作后果的喜好程度表示。由于任何动作序列都可以是一个规划,所以可能有很多规划。为了将这两种规划方法结合起来,我们必须在决策论框架中定义目标。我们说机构有一个目标 g 当且仅当:(1)机构在所有的动作中偏好动作 a_i 若 a_i 的执行将世界带到一个跟目标状态(包含 g 的状态)相对最近的可能状态;(2)由这样的状态组成的序列的执行将世界由初始状态带到目标状态。简言之,机构偏好这样的动作序列,其输出使 g 为真。

因果关系在人工智能规划中起着重要的作用。但是在决策论规划中,世界状态和动作之间不允许有因果关系。这有可能导致机构的非理性的决策,也不能通过规划的执行将世界由初始状态带到目标状态。因此,必须建立一个动作影响世界状态的因果关系模型。为此,我们首先将规划问题描述如下:一个目标 g ,一个状态集合 $S=\{s_1, \dots, s_n\}$,一个初始状态概率分布; $p(s_1), \dots, p(s_n)$;一个动作集合 $A=\{a_1, \dots, a_m\}$ 。

在状态集 S 中,目标状态记为 s_g 。因为任何包含 g 的状态都是目标状态,所以,目标状态可能不止一个。为了区分规划过程的不同阶段,我们给动作和状态加上上标。一个规划中的第 k 个动作记作 a^k ($k=0, 1, \dots$ 且

$a^k \in A$),在 a^k 执行时的状态概率分布为 $p(s_1^k), \dots, p(s_n^k)$, ($s_1^k=s_1, \dots, s_n^k=s_n, k=0, 1, \dots$)。

因果模型要回答的一个根本问题是动作 a^k 执行后新的状态概率分布是什么。假定动作 a^k 执行时世界在状态 s_j^k , a^k 执行后世界可能处于新状态 s_i^{k+1} 中 ($i, j=1, \dots, n$)。显然, s_i^{k+1} 出现的概率是以 s_j^k 和 a^k 为条件的概率,即 $p(s_i^{k+1} | s_j^k, a^k)$ 。规划机构除了 a^k 执行时的状态分布外没有关于世界的完全的知识。 a^k 执行后的状态概率分布可以按下式计算:

$$p(s_i^{k+1}) = \sum_{j=1}^n p(s_i^{k+1} | s_j^k, a^k) p(s_j^k, a^k) \quad (10)$$

此处 $i=1, \dots, n$ 。因为 s_j^k 和 a^k 无依赖关系且 a^k 肯定要执行,所以, $p(s_j^k, a^k) = p(s_j^k)$ 。因此,我们有

$$p(s_i^{k+1}) = \sum_{j=1}^n p(s_i^{k+1} | s_j^k, a^k) p(s_j^k) \quad (11)$$

此处 $i=1, \dots, n$ 。

这个因果关系模型表明动作通过其执行改变状态概率分布来影响世界。一旦新的状态概率分布已知,极大化主期望效用规则就可以用来选择下一个动作。注意,我们的因果模型不考虑“征兆性”动作,即那些仅仅是状态出现的证据而不真的造成这些状态的那些动作。

动作执行后,机构并不确切知道世界处于什么状态中。于是产生这样这一个问题:机构应在何时停止规划?在确定性规划中,一旦目标为真规划即停止。在非确定性规划中,规划机构仅仅知道目标状态的概率。令 p_g 为目标状态概率。若 p_g 大于等于预定的门槛数 α ($0 \leq \alpha < 1$),则停止规划,否则继续规划。若继续规划不能增大 p_g ,也停止规划。 α 是一个与具体领域有关的数,由专家确定。

对上面描述的规划问题,其初始状态概率分布为 $p(s_1^0), \dots, p(s_n^0)$,我们有如下的基本规划步骤:

- (1) 确定目标状态 s_g 并选取一个门槛数 α ;

(转第26页)

就是说, $L(a)$, $\neg E(a)$ 是被(强)认可的, $\neg L(a)$, $E(a)$ 是被拒斥的。

这说明, SH的选择机制倾向于认为这个人是律师。

在这个例子里, 关键之处在于“先验”硬度分配。即把身份特征描述置于比样本中的比例更“硬”的地位。一旦身份特征描述符合了律师这一身份的要求, 在样本中的比例这一因素就下降到次要地位, 这与我们的直观是非常符合的。顺便指出, 用现有非单调逻辑来处理这一问题, 无法摆脱“多扩张”问题的困扰。而用现有不确定推理方法来处理这一问题, 也因缺乏连贯的理论解释而陷于悖论。在SH中, 有确定因素与非单调因素通过带硬度标记的推理统一起来了, 这一问题也因之获得了简单而圆满的解答。

四、结 束 语

本文是一项很初步的工作。基于辩论的选择机制是在证明论的范围内叙述的, 它的模型论语义基

础的工作尚未完成。此外, 这里处理的主要是表示和推理, 关于信念及其评价的获取问题是极其重要的, 但本文已无法涉及。本文中 Γ 里各公式的评价部分显得似乎带有“先验”性, 实际上, 它们都应是学习的结果, 反馈的结果。把整个问题放到一个大循环当中来看, 硬度是可以动态升降的, 而受到“例外”攻击的历史记录, 是决定硬度升降的关键。作者觉得, 如能把获取、表示和推理完整地连成一体, 必将在对信念的认识上产生一个很大的飞跃。

最后, 作者想说明, 认知硬度不是也不应该是对信念的唯一评价尺度。非单调逻辑界最近十分关注的“Specificity”^[12]很可能是另一个重要的评价指标。如果能在SH中加进Specificity方面的因素, 必将会覆盖更多、更有意义的一类非单调现象。作者已在这个方向上有初步的进展, 具体结果将在另文中给出。

感谢李未教授、林作铨先生、王献昌博士的启发和帮助。感谢曲非同志为本文所提的建设性意见。(参考文献略)

(接第60页)

(2) 用公式(8)计算 μ_{ij} ($i=1, \dots, m, j=1, \dots, n$);

(3) 置 k 为0;

(4) 若 $p_k \geq \alpha$, 停止规划;

(5) 用公式(9)计算 $SEU(a_i)$ ($i=1, \dots, m$), 选取对应于 $SEU(a_i)$ ($i=1, \dots, m$)的最大值的动作作为 a^k ;

(6) 用公式(11)计算新的状态概率分布 $p(s_j^{k+1})$, $j=1, \dots, n$;

(7) k 增加1;

(8) 转向步骤(4)。

上述步骤产生的规划为 $a^0 a^1 a^2 \dots a^k \dots$, 且

$a^k \in A$ ($k=0, 1, \dots$)。若开始时 $p_0 \geq \alpha$, 则不需规划。

五、结束语

本文提出了一种传统人工智能和决策论相结合的非确定性规划方法。在这种方法中, 两种传统的规划方法步骤是相互交叉的。要使这种方法实际可行, 必须在以下两方面进一步开展研究。第一, 构造好的效用函数以便加快规划进程。第二, 构造符合实际的动作-状态因果关系模型。这两项工作都密切依赖于具体的应用领域。在非确定性规划中, 规划终止问题非常重要。好的效用函数有助于但不能保证规划终止。(参考文献略)