基于限定的溯因问题求解*

陈保平　孙吉贵

（吉林大学计算机系　长春　130023）

（吉林大学符号计算与知识工程开放实验室　长春　130023）

摘要　溯因问题是人工智能中的一个重要研究方向，它在许多领域中有着广泛的应用，但在很多情形下，溯因解释的求取是非常困难的. 本文提供一种基于限定理论的溯因解释求法，对于满足完备性公理和正原因假设的理论，可以证明其限定中的相容解释就是溯因解释，并且对Horn子句集给出具体的求解算法.

关键词　溯因推理，溯因解，限定，相容解释，独立的溯因解.

中图分类号　TP18

溯因问题是这样一类问题，已知一个原因集，一个结果集，一个背景理论（原因和结果之间的关系）. 我们观察到一些结果，寻找产生这些结果的可能原因.

溯因推理是人工智能中非常重要的研究课题，有着广泛的应用前景. 例如，在医疗诊断系统中，根据病人的症状和其它观察，形成一个最终的结论，这是一个溯因问题. 在自然语言理解领域，溯因解提供为什么这句话会被说出来的一个可行的解释.[1]在机器学习领域，溯因解被用于寻找知识库中与当前知识相关的部分，以便修改知识库中的错误和增添新的知识[2]等等.

通常情况下，溯因问题的求解是非常困难的，比如对某一观察，溯因解往往会有很多，其中许多是不一致的，确定或选择有效的且一致的溯因解一般是NP难度的. 在有些领域中，许多概念是采用相容性来定义的，如诊断等，我们称之为相容解释. 下面我们可以看到，在一定情形下，相容解释就是溯因解释.

本文把McCarthy的限定理论同溯因结合起来，在限定中研究溯因问题，对于一类广泛的溯因问题（满足本文中的完备性公理和正原因有效两条假设），我们证明了限定中的相容解释就是原溯因问题的溯因解，从而提供了求解一类溯因问题的理论与方法. 特别对于背景理论是Horn子句集的问题，本文还给出了计算溯因解的具体算法.

1　溯因问题与相容解释

我们首先定义溯因问题.

* 本文研究得到国家自然科学基金、国家863高科技项目基金和攀登计划基金资助. 作者陈保平，1971生，硕士生，主要研究领域为定理机器证明与自动推理. 孙吉贵，1962年生，副教授，主要研究领域为定理机器证明与自动推理.

本文通讯联系人：陈保平，长春130023，吉林大学计算机系

本文1996-04-08收到修改稿

定义 1.[3]　一个溯因问题是三元组$\langle C,E,T\rangle$,其中 C 是原子集,叫作原因集;E 是原子集叫作结果集;T 是任意子句集,叫作背景理论.一个观察 $g(g\in E)$的溯因解 A 定义如下:

(1) $A\subseteq C$,且 A 与 C 一致;

(2)$(T\cup A)\vdash g$;

(3) A 是满足上述条件在集合包含关系下的极小集合.

定义 2. $\langle C,E,T\rangle$定义同上,观察 g 是 E 的元素,关于 g 的否定集 D 定义如下:

(1) $D\subseteq C$　　(2) $T\cup\{g\}\cup\{\sim d\mid d\in D\}$一致.

当否定集在集合包含关系下极大时,我们有 $T\cup\{g\}\cup\{\sim d\mid d\in D\}\vdash C-D$ 中文字的合取式,此时称 $C-D$ 叫作关于 g 的相容解释.

Reiter 把诊断问题定义成三元组$\langle SD,C,OBS\rangle$[1],其中 SD 和 OBS 是一阶公式集,SD 是系统描述,OBS 是观察集,C 是谓词 AB 的集合,AB 代表异常.一个诊断定义成满足下式的最小的$\triangle$,$\triangle\subseteq C$,且 $SD\cup OBS\cup\{c\mid c\in\triangle\}\cup\{\sim c\mid c\in C\backslash\triangle\}$一致.可以看出,一个诊断就是一个相容解释.

溯因解与相容解释不同,对于溯因解释,要求在背景理论中可以推导出观察.而对于相容解释,则是在无矛盾的前提下,尽可能的否定原因集中的文字.作为文字的集合,一个溯因解释在通常情况下是某一相容解释的子集.

2　限定中的溯因解

限定推理是由 McCarthy 提出的一种强有力的非单调推理,有关限定的定义见文献[4].通常,限定的定义包含二阶量词,所以是二阶公式.二阶公式难于表达,也难以使用,但是在一些情形下限定可以转化为等价的一阶公式,可使用各种对一阶公式的操作来处理.关于限定的转换见文献[5],这里我们仅叙述关于限定的 2 个基本结果.

定理 1. M 是 $CIRC(T;P;Z)$的模型当且仅当 M 是 T 的(P,Z)最小模型,也就是说对任意公式 F,$CIRC(T;P;Z)\models F$ 当且仅当对每个 T 的(P,Z)最小模型 M,$M\models F$.

定理 2. 对任意公式 F,F 由 $P^+\vee Q^+\vee Q^-$ 中谓词组成,则 $T\vdash F$ 当且仅当 $CIRC(T;P;Z)\models F$,其中 Q 是 T 中除 P,Z 以外的谓词集合,P^+,O^+代表 P,Q 中谓词的正出现,Q^- 代表 Q 中谓词的负出现.

下面我们考虑溯因问题$\langle C,E,T\rangle$,并做如下假设:

(1)完备性公理:设 $A_1,A_2,...,A_n$ 是 g 在 T 中的所有溯因解,对于任意公式 S,若 $S\vdash g$在 T 中成立,则存在 A_i,使 $A_i\vdash S$ 在 T 中成立,即 g 在 T 中成立当且仅当存在某个 A_i 在 T 成立.

(2)正原因有效:本文要求可溯因集(即原因集 C)是原子集,且任意 C 中原子的负出现是无效的.也就是说,对任意 $c_1,c_2,...,c_n,\sim b_1,...,\sim b_m\vdash g$,$c_i,b_j$ 是 C 中原子,则 $c_1,c_2,...,c_n\vdash g$.

完备性公理是指对任意观察结果,我们总可以在原因集中得到解释;正原因有效是指一些观察是因为某些原因发生而产生的,而不是由于某些原因不发生而产生的,因此上述 2 个假设是合理的.

令 Q 为 T 中除 C,E 中谓词以外的谓词集合，则我们考虑 $\langle C,E,T\rangle$ 与限定 $CIRC(T;E;Q)$ 之间的关系.

引理 1. T 不一致，当且仅当 $CIRC(T;E;Q)$ 不一致.

证明：设 T 不一致，由限定的定义，$CIRC(T;E;Q)$ 不一致. 设 $CIRC(T;P;Q)$ 不一致，由定理 1，所有最小模型为假，再由最小模型定义，T 不一致.

引理 2. 设 T 一致，$A\subseteq C$，则 $A\cup T$ 一致当且仅当 $A\cup CIRC(T;E;Q)$ 一致.

证明：设 $T\cup A$ 一致，由 T 一致及引理 1，可知 $CIRC(T;E;Q)$ 一致；如果 $A\cup CIRC(T;E;Q)$ 不一致，则 $CIRC(T;E;Q)\models\sim A$，由定理 2，$T\vdash\sim A$，与 $A\cup T$ 一致矛盾. 同理可证充分性.

定理 3. A 是 g 在 T 中的溯因解，当且仅当 A 是 g 在 $CIRC(T;E;Q)$ 中的溯因解.

证明：设 A 是 g 在 T 中的溯因解，则 $A\cup T$ 一致，由引理 2 可知 $A\cup CIRC(T;E;Q)$ 一致. 又因为 $A\cup T\vdash g$，即 $T\models(A\rightarrow g)$，据定理 2 得 $CIRC(T;E;Q)\models A\rightarrow g$，所以 $A\cup CIRC(T;E;Q)\models g$. 故 A 是 $CIRC(T;E;Q)$ 的溯因解. 同理可证定理的充分性.

定义 3. $A_1,A_2,\ldots,A_n$ 是 g 在 T 中的所有溯因解，若 $A_i\cup T\nvdash A_j(j\neq i)$，则称 A_i 在 T 中是独立的.

定义 4. $A_1,A_2,\ldots,A_n$ 是 g 在 T 中的所有溯因解，则 $A=A_1\vee A_2\vee\ldots\vee A_n$ 叫作 g 的全解释.

定理 4. 设溯因问题 $\langle T,C,E\rangle$，其中 C,E 是原子集，Q 定义同上，满足完备性公理，则 $CIRC(T;E;Q)\equiv T\cup\{g_i=A_i\}$，其中 $E=\{g_1,g_2,\ldots,g_n\}$，A_i 是 g_i 的全解释.

证明：设 $B_1,B_2,\ldots,B_m$ 是 T 中所有满足 $B\rightarrow g_i$ 的公式，将 $\vee B_i(i=1,\ldots,m)$ 化为析取范式，变为 $\vee D_i$. 令 $A_1,A_2,\ldots,A_n$ 是 g_i 在 T 中的所有溯因解，由完备性公理，对任意的 D_i 存在 A_j 使 $A_j\rightarrow D_i$，则 $(\vee A_i)\rightarrow(\vee D_j)$.

设 $P'=(\vee A_i)$，则 $P'\rightarrow g_i$. 由限定的定义，在 $CIRC(T;E;Q)$ 中 $P'=g_i$，P' 是 g_i 在 T 中的全解释. 对于所有的 g_i，存在相应的全解释 P'，使 $P'=g_i$ 成立. 故定理成立.

引理 3. 设 A 是 g 在 T 中的一个溯因解，则 A 在 T 中独立当且仅当 A 在 $CIRC(T;E;Q)$ 中独立.

证明：此引理可由定理 2 和定理 4 直接得出.

引理 4. $CIRC(T;E;Q)\cup\{g\}$ 不一致当且仅当 g 在 T 中无溯因解.

证明：由定理 2 及引理 1 得出.

定理 5. 设溯因问题 $\langle T,E,Q\rangle$，T,C,E,Q 要求同上，满足上面的 2 个假设；若 $T\nvdash(\vee c_i)$（其中 c_i 是 C 中的文字），则 A 是 g 在 T 中的一个独立的溯因解当且仅当 A 是 g 在 $CIRC(T;E;Q)$ 中的一个相容解释.

证明：(1) 设 A 是 g 在 T 中的一个独立的溯因解，则由引理 3 和定理 3 得，A 是 g 在 $CIRC(T;E;Q)$ 中独立的溯因解释. 若在 $CIRC(T;E;Q)$ 中，$g=A\vee B$，B 是其余的溯因解的析取，由 A 独立知 $A\cup CIRC(T;E;Q)\nvdash B$. 又由 $A\cup T$ 一致，据引理 2，$A\cup CIRC(T;E;Q)$ 一致，故 $A\cup CIRC(T;E;Q)\cup\sim B$ 一致.

注意到 $A\cup CIRC(T;E;Q)\models g$，所以 $CIRC(T;E;Q)\cup\{g\}\cup\sim B$ 一致. 令 M 是 $CIRC(T;E;Q)\cup\{g\}\cup\sim B$ 的一个模型，$D=\{a_1,a_2,\ldots,a_n\}$ 是从 B 的每个溯因解中取出在 M 中

为假的文字组成的集合,则$\{g\}\cup\sim D$一致,且$CIRC(T;E;Q)\cup\{g\}\cup\sim D\vdash A$,于是$C-D$是一个相容解释,且$A\subseteq C-D$.

设$R=C-A$,由假设$T\nvdash(\vee c_i)$,c_i是R中文字.依据定理2得$CIRC(T;E;Q)\nvdash(\vee c_i)$,从而$CIRC(T;E;Q)\cup\sim R$一致,且$CIRC(T;E;Q)\cup\sim R\nvdash\sim g$,故$CIRC(T;E;Q)\cup\{g\}\cup\sim R$一致.显然$R$已经极大,$C-R=A$是相容解释.因此任意独立的溯因解都是相容解释.

(2)设$C-D$是一个相容解释,则$CIRC(T;E;Q)\cup\{g\}\cup\sim D$一致,下面我们用反证法证明$C-D$必包含某一溯因解.假设$C-D$不包含任意溯因解,那么对于任意溯因解$A_i$,至少存在$A_i$中的一个文字在$D$中,于是$\sim D\rightarrow\sim A_i$,故$\sim D\rightarrow\sim(\vee A_i)$.由于$g=(\vee A_i)$,所以$\sim D\rightarrow\sim g$,与$CIRC(T;E;Q)\cup\{g\}\cup\sim D$一致矛盾.从而$C-D$必包含某溯因解$A$.

注意到$T\nvdash(\vee L_i)$,(L_i是C中文字),A是T中的独立的溯因解.则由(1)的证明知,A必是一个相容解释,由$A\subseteq C-D$得$C-D=A$,即任意相容解释必是一个溯因解.定理得证.

从定理5可以看出,象诊断这类用相容性定义的概念在限定中与溯因解释等价,可以用诊断的方法求解溯因解释,也可以用溯因的方法求取诊断.上面的定理除要求满足2个假设外,还要求$T\nvdash(\vee c_i)$,c_i是C中文字,即C中原子不能互相推导.

下面我们举一个简单的例子.

我们对一间屋子进行观察,结果是*house_dark_quiet*,屋子的描述用理论T表示如下:

house_dark ∧ *house_quiet*→*house_dark_quiet*

light_out→*house_dark*

no_one_home ∨ *blackout*→*light_out*

tv_off ∧ *radio_off*→*house_quiet*

no_one_home ∨ *no_shows* ∨ *blackout*→*tv_off*

no_one_home ∨ *bad_songs* ∨ *blackout*→*radio_off*

令C={*no_one_home*, *blackout*, *no_shows*, *bad_songs*},E={*house_dark_quiet*},Q={*light_out*, *tv_off*, *radio_off*, *house_dark*, *house_quiet*}.

则$CIRC(T;E;Q)=T\cup$\{*house_dark_quiet*=(*no_one_home* ∨ *blackout*)\}.

$CIRC(T;E;Q)$的相容解释有*no_one_home*和*blackout*,也就是*house_dark_quiet*在T中的溯因解.

3 Horn 子句集的溯因解算法

定义5[6]. 对于Horn子句集,若每个子句恰含一个正文字,则称此Horn子句集为Definite子句集.

下面用我们上节给出的方法对Definite子句形式的溯因问题计算其溯因解.设$\langle T,C,E\rangle$是溯因问题,T是Definite子句集,C,E是原子集.且C中原子只出现在子句体中,E中原子只出现在子句头中.

定义6.[6] T称为非循环的,如果存在从T中原子到正整数的映射*Level*,使下列条件成立:

为假的文字组成的集合，则 $\{g\}\cup\sim D$ 一致，且 $CIRC(T;E;Q)\cup\{g\}\cup\sim D\vdash A$，于是 $C-D$ 是一个相容解释，且 $A\subseteq C-D$.

设 $R=C-A$，由假设 $T\nvdash(\vee c_i)$，c_i 是 R 中文字. 依据定理 2 得 $CIRC(T;E;Q)\nvdash(\vee c_i)$，从而 $CIRC(T;E;Q)\cup\sim R$ 一致，且 $CIRC(T;E;Q)\cup\sim R\nvdash\sim g$，故 $CIRC(T;E;Q)\cup\{g\}\cup\sim R$ 一致. 显然 R 已经极大，$C-R=A$ 是相容解释. 因此任意独立的溯因解都是相容解释.

(2) 设 $C-D$ 是一个相容解释，则 $CIRC(T;E;Q)\cup\{g\}\cup\sim D$ 一致，下面我们用反证法证明 $C-D$ 必包含某一溯因解. 假设 $C-D$ 不包含任意溯因解，那么对于任意溯因解 A_i，至少存在 A_i 中的一个文字在 D 中，于是 $\sim D\rightarrow\sim A_i$，故 $\sim D\rightarrow\sim(\vee A_i)$. 由于 $g=(\vee A_i)$，所以 $\sim D\rightarrow\sim g$，与 $CIRC(T;E;Q)\cup\{g\}\cup\sim D$ 一致矛盾. 从而 $C-D$ 必包含某溯因解 A.

注意到 $T\nvdash(\vee L_i)$，(L_i 是 C 中文字)，A 是 T 中的独立的溯因解. 则由(1)的证明知，A 必是一个相容解释，由 $A\subseteq C-D$ 得 $C-D=A$，即任意相容解释必是一个溯因解. 定理得证.

从定理 5 可以看出，象诊断这类用相容性定义的概念在限定中与溯因解释等价，可以用诊断的方法求解溯因解释，也可以用溯因的方法求取诊断. 上面的定理除要求满足 2 个假设外，还要求 $T\nvdash(\vee c_i)$，c_i 是 C 中文字，即 C 中原子不能互相推导.

下面我们举一个简单的例子.

我们对一间屋子进行观察，结果是 *house_dark_quiet*，屋子的描述用理论 T 表示如下：

house_dark ∧ *house_quiet*→*house_dark_quiet*

light_out→*house_dark*

no_one_home ∨ *blackout*→*light_out*

tv_off ∧ *radio_off*→*house_quiet*

no_one_home ∨ *no_shows* ∨ *blackout*→*tv_off*

no_one_home ∨ *bad_songs* ∨ *blackout*→*radio_off*

令 C = {*no_one_home*, *blackout*, *no_shows*, *bad_songs*}，E = {*house_dark_quiet*}，Q = {*light_out*, *tv_off*, *radio_off*, *house_dark*, *house_quiet*}.

则 $CIRC(T;E;Q)=T\cup$ {*house_dark_quiet* = (*no_one_home* ∨ *blackout*)}.

$CIRC(T;E;Q)$ 的相容解释有 *no_one_home* 和 *blackout*，也就是 *house_dark_quiet* 在 T 中的溯因解.

3 Horn 子句集的溯因解算法

定义 5[6]. 对于 Horn 子句集，若每个子句恰含一个正文字，则称此 Horn 子句集为 Definite 子句集.

下面用我们上节给出的方法对 Definite 子句形式的溯因问题计算其溯因解. 设 $\langle T,C,E\rangle$ 是溯因问题，T 是 Definite 子句集，C，E 是原子集. 且 C 中原子只出现在子句体中，E 中原子只出现在子句头中.

定义 6.[6] T 称为非循环的，如果存在从 T 中原子到正整数的映射 *Level*，使下列条件成立：

(1) $level(L)=0$,L 是 C 中原子.

(2) 对于 T 中任意子句 $a_1,a_2,...,a_n \to b$,满足 $Level(b)>\max\{Level(a_i)\}$ $i=1,2,...,n$.

对于 $\langle T,C,E\rangle$ 同上所述,若 T 是非循环的命题 Definite 子句集,我们可以使用文献[5]中的理论和方法来具体计算限定 $CIRC(T;E;Q)$.

对 E,Q 中任一谓词 p,设 $B_1\to p,B_2\to p,...,B_n\to p$ 是 T 中所有以 p 为头的子句,B_i 是新的谓词名,代表文字 $a_{i1},...,a_{im}$ 的合取,$i=1,2,...,n$,则向 T 中加入子句 $B_i\to a_{i1}\wedge...\wedge a_{im}$,$p\to B_1\vee B_2\vee...\vee B_n$,对 E,Q 中所有谓词完成上述操作所得的子句集 T' 就是 $CIRC(T;E;Q)$.

由定理5和引理4可得到如下算法.

算法1. (计算 T 的溯因解)

(1) 计算 $CIRC(T;E;Q)$.

(2) 若 $CIRC(T;E;Q)\cup\{g\}$ 不一致,则无解.

(3) 否则,计算 $CIRC(T;E;Q)$ 的相容解释.其相容解释就是溯因解.

算法2. (求解 $CIRC(T;E;Q)$ 的相容解释)

```
D:={ }; N:={ }; S:=C;
while (S 不为空)
{从 S 中选取文字 L;
  S:=S-L;
  if {~L}∪CIRC(T;E;Q)∪{g}一致
    then D:=D∪{L}
    else N:=N∪{L};
}
```

此算法结束后,N 就是 $CIRC(T;E;Q)$ 的相容解释,也就是 T 的溯因解.

4 结 语

本文提供了一种基于限定理论的溯因解求法,对于满足完备性公理和正原因假设的理论,证明了其限定中的溯因解就是原理论的溯因解,而限定中的溯因解就是相容解释,从而可以通过求取限定中的相容解释来得到溯因解.特别的,对于 Horn 子句集,其限定很容易求得,我们给出一个算法来求取溯因解.对于诊断等使用相容性来定义的概念,我们找到了它们与溯因解释之间的关系,从而为求解溯因解释提供了新的方法.

参考文献

1 Bylander T. The computational complexity of abduction. Artificial Intelligence, 1991,**49**:25~60.

2 Raedt L D. Belief updating from integrity constraints and queries. Artificial Intelligence, 1992,**53**:291~307.

3 Konolige. Abduction versus closure in casual theory. Artificial Intelligence, 1992,**53**:255~289.

4 Gelfond M. On the relationship between circumscription and negation as failure. Artificial Intelligence, 1989,**38**:89~126.

5 Lifschiz V. Computing circumscription. In: Aravind J ed. Proceedings of IJCAI_85, U.S., 1985. 121~127.

6 Cavedon L. Acyclic logic program and the completeness of SLDNF-resolution. Theoretical Computer Science,

1991,**86**:81～92.

ABDUCTIVE REASONING WITH CIRCUMSCRIPTION

CHEN Baoping SUN Jigui

(*Department of Computer Science Jilin University Changchun* 130023)

(*Open Laboratory for Symbol Computing and Knowledge Engineering Jilin University Changchun* 130023)

Abstract Abduction is an important research work in AI(artificial intelligence), and can be applied in many other research fields. But it is difficult to compute abductive explanations. In this paper, a method used in circumscription is given. When a theory is suited with complete axiom and positive reason assumption, it can be proved that the consistent explanation is exactly the abductive explanation in the circumscription. Furthermore, an algorithm using in Horn clauses is given.

Key words Abduction, abductive explanation, circumscription, consistent explanation, independent abductive explanation.

Class number TP18