

基于优先解释的不完全信息推理及其应用*

叶风 徐晓飞 王亚东

(哈尔滨工业大学计算机科学与工程系 哈尔滨 150001)

摘要 不完全信息下的近似推理是知识工程面临的困难问题之一. 文章提出了一种具有非单调性质的优先逻辑程序理论. 该理论能够对知识的解释进行综合评判, 进而优选解释, 使其成为现有知识的最佳理论逼近, 达到在择优意义下的理论完全化, 避免了对知识的完全性及一致性要求. 为获取应用领域的优先逻辑程序, 基于归纳逻辑程序设计技术设计了一种多方法归纳学习算法, 该算法具有较强的归纳能力. 此理论与算法已应用在 863 农业专家系统中, 并获得满意结果.

关键词 专家系统, 不完全信息推理, 优先解释, 归纳逻辑程序设计.

中图法分类号 TP18

不完全信息下的近似推理是知识工程面临的最困难问题之一, 也是人工智能研究中的重要课题, 因而有着重要的理论与应用价值. 人工智能基础领域对人类在不完全信息下, 从含有错误甚至矛盾的信念中自如地进行推理这一机制进行了大量研究, 诸如常识推理的研究. 这类研究工作导致一批刻画这一机制的形式系统^[1~3]的出现, 这些系统均有非单调性质. 这些理论总的特征是刻画在某些特定信息未知情况下, 系统应作出的各种信念并以此扩充知识, 此类理论面临的最大问题是, 其计算难解性与理论修正困难. 模拟这一机制的另一方案是连接系统, 神经网络具有快速计算及容错的能力, 是对付不完全信息的有力工具. 然而, 神经网络缺乏足够的表达能力以处理高级推理活动.

G. Pinkas 的罚逻辑^[4]是命题级上的工作, 仅是矛盾命题的一种解决方案. 本文在此基础上提出了一种基于正规逻辑程序的优先理论, 将连接网络中的权值思想引入逻辑程序中, 从而得到优先逻辑程序子句理论(简称优先理论). 这一理论是一阶的, 并自然地具有逻辑程序计算上的优越性, 利用这一理论可以得到所表示的知识的优先解释. 优先解释是依据优先理论进行择优的结果, 是标准逻辑程序模型的自然推广, 表达了一种与现有工作完全不同的推理机制. 研究表明, 优先解释这一新的推理机制能够很好地刻画人类在不完全信息下的直觉推理, 能够有效地处理矛盾或不完全知识的推理问题. 本文的另一项工作是, 设计一个基于归纳逻辑程序设计^[5,6]方法 ILP(inductive logic programming)的学习算法 BDI(bi-directional induction), 利用这一算法能够归纳产生应用领域的优先理论. 最后通过一个专家系统实例, 说明本文理论与算法的应用情况.

1 优先逻辑程序理论

在不完全信息下进行推理的关键是使知识相对完全化. 传统做法是将某种信念施加于整个知识集来扩充知识, 这种做法没有考虑各知识个体的可靠性对整体所产生的影响, 因此存在各种问题. 本节定义一种有关个体知识可靠性及知识解释评价的理论.

1.1 优先逻辑程序语法

表达可靠性常用的方式有概率、主观信度等. 本文采用 Pinkas^[4]的处理方式, 即赋予知识实体一个实数, 对

* 本文研究得到国家 863 高科技项目基金资助. 作者叶风, 1960 年生, 博士生, 主要研究领域为机器学习, 人工智能逻辑基础. 徐晓飞, 1962 年生, 博士, 教授, 博士生导师, 主要研究领域为计算机集成制造, 分布式数据库. 王亚东, 1964 年生, 副教授, 主要研究领域为机器学习, 专家系统.

本文通讯联系人: 叶风, 哈尔滨 150001, 哈尔滨工业大学计算机科学与工程系

本文 1998-01-19 收到原稿, 1998-03-23 收到修改稿

其不作特别解释,以表达一种一般框架,这使得许多系统能够轻易地纳入其中.

下面的优先逻辑程序部分地采用了 Pinkas 的罚逻辑(penalty logic)处理方式,赋予每一程序子句一个实数,这使通过子句得到的结论有了度量.

定义 1. 优先逻辑程序是逻辑程序的扩充,是形如 $r: P \leftarrow Q_1, \dots, Q_n$ 的扩充子句集. 其中 $r \in R^+$ 是正实数,称为优先值; P 是原子,称为子句头; Q_i 是文字,称为子句体(子目标).

为表述方便,常将扩充子句记为集合形式 $\{P, \neg Q_1, \dots, \neg Q_n\}$.

优先逻辑程序的推理过程与通常逻辑程序的差别仅在于,若结论 p 由 $r: p \leftarrow Q_1, \dots, Q_n$ 得到,则赋予 p 实数 r , 记为 $r: p$, 意为 p 得到的支持程度为 r .

优先逻辑程序提供了一种在冲突与多重命题间作出选择的机制,这即是这一理论的意义所在. 为此,考虑 Reiter 在讨论非单调推理时给出的一个著名例子.

例 1: 考察 Nixon 的政治倾向. 如果已知任何人或为共和党人或为基督徒, 共和党人都不是和平主义者, 基督徒都是和平主义者, 问: Nixon 是和平主义者还是非和平主义者? 以上事实的优先逻辑程序为

- 1000: Person(Nixon).
- 20: Republican(X) \leftarrow Person(X).
- 20: Christian(X) \leftarrow Person(X).
- 10: Not_Pacifist(X) \leftarrow Republican(X).
- 10: Pacifist(X) \leftarrow Christian(X).

上述子句中的优先值都是假定的. 从中得到 10: Pacifist(Nixon) 与 10: Not_Pacifist(Nixon) 两个优先值相等而内在语义冲突的结论. 因为, 一般认为基督徒是和平主义者, 而共和党人反之. 导致这种反常结论的直接原因是规则 10: Not_Pacifist(X) \leftarrow Republican(X) 与 10: Pacifist(X) \leftarrow Christian(X) 具有平等的重要性. 如果后者的优先值改为 15, 则两种结论的优先量不再相等, 结论 15: Pacifist(Nixon) 相对优于 10: Not_Pacifist(Nixon), 即 Pacifist(Nixon) 得到更多的支持.

例 1 说明, 优先逻辑程序以优先值方式丰富了逻辑程序的表达能力, 下面还将看到这一简单的处理方式对不完全信息的处理也是适宜的.

1.2 优先逻辑程序的解释

利用优先值可对逻辑程序的解释进行评判. 一般地, 事实具有较大优先值, 表示其相对得到较多的支持, 规则具有较大优先值, 反映了这一规则具有较大可靠性. 将解释的优先级定义为其中各基本事实优先值的和. 如此, 解释的优先级大小反映了知识整体对解释的支持程度. 优先逻辑程序的优先解释倾向于选择那些得到较多支持的基本事实.

本文只讨论非递归逻辑程序, 为了保证有可计算性, 也需假定知识中无函数存在(有限 Herbrand 域). 限于篇幅, 略去定理证明.

以下理论或术语除优先值外均为经典的. 令 Ψ 为优先逻辑程序, $U_H(\Psi)$ 表示 Ψ 的 Herbrand 域, $B_H(\Psi)$ 表示 Ψ 的 Herbrand 基, $Inst(\Psi)$ 表示 Ψ 在 $U_H(\Psi)$ 域上例化子句的全体, $M_H(\Psi)$ 表示 Ψ 的 Herbrand 模型的正原子部分. Ψ 的基文字集 I 记为 $I = I^+ \cup \neg I^-$, 其中 $I^+ = \{e | e \in B_H(\Psi) \text{ 且 } e \in I\}$, $I^- = \{e | e \in B_H(\Psi) \text{ 且 } \neg e \in I\}$.

定义 2. Ψ 的基文字集 I 称为 Ψ 的解释当且仅当 $I^+ \cap I^- = \emptyset$.

优先逻辑程序 Ψ 的解释 I 的优先值计算函数 $Prec$ 定义如下.

定义 3. $Prec(I, \Psi) = \sum_{e \in I^+ \wedge (r: e \leftarrow Q_1, \dots, Q_n \in Inst(\Psi) \wedge \forall i, Q_i \in I)} r + \sum_{e \in I^- \wedge \exists (r: e \leftarrow Q_1, \dots, Q_n \in Inst(\Psi) \wedge \forall i, Q_i \in I)} \lambda$, 其中 I 为 Ψ 的解释.

$Prec$ 计算 I 中每个基文字在 Ψ 中得到的优先值, 再将这些优先值取和得到 I 的优先值. 常数 λ 表示当一属于解释 I 的基本事实在 Ψ 中未得到任何支持时, 系统对其“惩罚”. $Prec$ 建立了解释与优先理论间的关系, 其计算值综合评价了解释对知识的适宜程度. 最好的解释即获得最高优先值.

令 $n = |Inst(\Psi)|$, Ψ 中子句的基数, $m = |I|$, $\mu = \text{Max}\{r | r: P \leftarrow Q_1, \dots, Q_n \in \Psi\}$.

性质 1. $\lambda \leq Prec(\{e\}, \Psi) \leq n\mu$.

性质 2. $\lambda m \leq Prec(I, \Psi) \leq \mu nm$.

定义 4. 解释 I 称为优先逻辑程序 Ψ 的优先解释当且仅当 $Prec(I, \Psi) = \text{Max}\{Prec(I, \Psi) \mid I \text{ 是 } \Psi \text{ 的解释}\}$.
令 $\Gamma(\Psi) = \{I \mid I \text{ 是 } \Psi \text{ 的优先解释}\}$.

定理 1. 若 $I \in \Gamma(\Psi)$, 则 $M_H(\Psi) \subseteq I$.

优先解释必覆盖标准模型的正原子部分, 因此, 优先解释是对标准模型的适当扩充.

定理 2. 若 $I \in \Gamma(\Psi), e \in I$, 则 $Prec(I - \{e\}, \Psi) \leq Prec(I, \Psi)$.

优先解释中每一事实都不会降低解释的优先级. 在优先解释框架下, 解释的稳定部分在实用中有重要意义.

定义 5. 集 K 称为 Ψ 的优先解释核, 如果 $K = \bigcap_{I \in \Gamma(\Psi)} I$.

推论 1. Ψ 的优先解释核 $K \supseteq M_H(\Psi)$.

例 2: 仍考虑例 1 事件. 去掉事实 1000: Person(Nixon), 增添 Not_Pacifist 与 Pacifist 具有相反语义这一事实, λ 取 -10, 论域为 {Nixon}.

- (1) 20: Republican(X) \leftarrow Person(X).
- (2) 20: Christian(X) \leftarrow Person(X).
- (3) 10: Not_Pacifist(X) \leftarrow Republican(X).
- (4) 10: Pacifist(X) \leftarrow Christian(X).
- (5) 100: Not_Pacifist(X) $\leftarrow \rightarrow$ Pacifist(X).
- (6) 100: Pacifist(X) $\leftarrow \rightarrow$ Not_Pacifist(X).

据 Prec 定义不难得到仅有的两个优先解释, $I_1 = \{\text{Person(Nixon), Republican(Nixon), Christian(Nixon), Pacifist(Nixon)}\}$, $I_2 = \{\text{Person(Nixon), Republican(Nixon), Christian(Nixon), Not_Pacifist(Nixon)}\}$, 优先值均为 140, 逻辑程序 Ψ 的优先解释核为 $\{\text{Person(Nixon), Republican(Nixon), Christian(Nixon)}\}$, 注意到这一稳定部分中的事实 Person(Nixon) 是猜出来的. 若第 4 子句的优先值改为 15, 则仅得到唯一优先解释 I_1 , 而当第 4 个子句的优先值改为 5 时, 又得到唯一优先解释 I_2 . 由此可见, 优先解释框架具有非单调性.

1.3 优先解释的构造

解释相对于优先理论的“优劣”, 取决于优先理论本身与系统罚值 λ , λ 值的大小决定了这一框架对解释选择的“保守”与“激进”. 由 Prec 的定义不难得知, λ 越小, 优先解释越接近逻辑程序的 Herbrand 模型, 即为“保守”方式; 反之, 优先解释含有越多的非 Herbrand 模型部分, 即为“激进”方式, λ 的取值应视具体应用而定.

性质 3. 若 $\lambda = -\infty$, 则 $\Gamma(\Psi) = \{M_H(\Psi)\}$.

理想情况是给出优先解释构造的有效算法, 但优先解释的构造在计算上是困难的. 然而在实际应用中, 我们只需考虑一个特定事实是否属于优先解释即可. 特别是对有定优先逻辑程序, 有下述定理保证算法的存在.

定义 6. 集 Ext 称为基原子 e 关于优先逻辑程序 Ψ 的扩张集, 如果

$$Ext = \{e\}.$$

$$Ext = T \cup \{Q_1, \dots, Q_n \mid e \leftarrow Q_1, \dots, Q_n \in \text{Inst}(\Psi)\}.$$
 其中 T 是 e 的关于 Ψ 的扩张集.

$$Ext = T \cup \{Q_1, \dots, Q_n \mid e \leftarrow Q_1, \dots, Q_{i-1}, e, Q_{i+1}, \dots, Q_n \in \text{Inst}(\Psi)\}.$$
 其中 T 是 e 的关于 Ψ 的扩张集.

定理 3. $e \in \bigcap_{I \in \Gamma(\Psi)} I$. 当且仅当存在 e 的关于有定优先逻辑程序 Ψ 的扩张集 Ext, 使得 $Prec(Ext, \Psi) > 0$.

由此, 判定是否 $e \in \bigcap_{I \in \Gamma(\Psi)} I$ 只需寻找 e 的扩张集 Ext, 使得 $Prec(Ext, \Psi) > 0$. 在优先有定程序下, 容易产生各扩张集.

2 学习算法

归纳逻辑程序设计 ILP 是机器学习领域新出现的一个研究分支, 其目标是从实例及背景知识中归纳产生目标的一阶描述, 特别是产生逻辑程序形式的概念描述. ILP 方法的主要特征是: (1) 在归纳学习过程中充分利用已有的背景知识; (2) 使用逻辑程序作为描述语言, 使概念描述精确. 作为知识获取手段, 本文通过设计基于 ILP 方法的学习算法来归纳产生与领域有关的优先逻辑程序, 并在渐进学习过程中使系统知识不断得以完善.

2.1 ILP 问题描述

给定背景知识 BK , 正反例集 $E = E^+ \cup E^-$, 其中 E^+ 代表正例集, E^- 代表反例集, ILP 方法归纳产生假设 H 并满足下述条件:

- ① 先验可满足性. $BK \wedge E^- \not\models \square$. ② 先验必要性. $BK \not\models E^+$.
 ③ 后验可满足性. $BK \wedge H \wedge E^- \not\models \square$. ④ 后验充分性. $BK \wedge H \models E^+$.

其中 \square 代表矛盾. ILP 方法的目的是, 在保证一致性前提下, 找出 H , 使得 BK 与 H 能够逻辑蕴涵 E^+ . 这里, BK, H, E^-, E^+ 约束在逻辑程序空间内.

归纳推理是演绎推理的逆过程, 因而归纳规则也多为演绎规则的逆. 因归结规则在演绎推理中是完备的, 使得逆归结^[5,6], 如 V 算子与 W 算子是 ILP 中经常采用的普化策略. 这些是从特殊到一般的普化规则. ILP 的另一种学习方法是使用从一般到特殊的特化规则. 混合使用普化与特化规则能够较充分地产生候选归纳目标.

2.2 学习算法

如何在逻辑程序空间中寻找最优描述子句是基于 ILP 方法学习算法的核心问题之一. 本文设计 ILP 学习算法 BDI, 该算法同时使用普化与特化规则产生假设子句 (优先逻辑程序子句), 并使用期望精度与加权精度增益函数来评估所得子句的质量. 以下凡实例均为形如 $P \leftarrow Q_1, \dots, Q_n$ 的基子句, P 是示例的类属谓词, Q_i 是相应的属性谓词. 称理论 T 覆盖示例 $P \leftarrow Q_1, \dots, Q_n$, iff $T, Q_1, \dots, Q_n \vdash P$. 反之, 称 T 排斥示例 $P \leftarrow Q_1, \dots, Q_n$. 如第 2.1 节中定义, 期望找到假设集 H , 使得对给定示例集 $E = E^+ \cup E^-, \forall e \in E^+, T$ 与 H 覆盖 $e, \forall e' \in E^-, T$ 与 H 排斥 e' .

期望精度与加权精度增益函数^[7]用来评价 $h \in H$ 的质量, 主要依据是 h 对 E^+ 的覆盖程度以及对 E^+ 与 E^- 的划分情况. 令 $n(h)$ 表示 h 覆盖的示例总数, $n^+(h)$ 表示 h 覆盖的正例总数, 则 h 的期望精度 $AC(h) = P(+/h)$, 即 h 覆盖的示例为正例的概率. 对多类学习问题, Laplace 估计是 $P(+/h)$ 的较好近似, $P(+/h) = \frac{n^+(h)+1}{n(h)+k}$, 其中 k 是 E 中示例的类数. 若假设 h' 是将归纳规则运用于假设 h 而得到的, 则 h' 相对于 h 的优劣可利用加权精度增益函数 WAC 来估计. h' 相对于 h 的加权精度增益 $WAC(h', h) = \frac{n^+(h')}{n^+(h)} (AC(h') - AC(h))$.

定义 7. 称子句 h 在 BK 与 E 环境下关于 H 是冗余的, 如果 $\forall e \in E^+, BK, H - \{h\}$ 覆盖 e .

对示例 $P \leftarrow Q_1, \dots, Q_n \in E, P$ 成为类属谓词 (规则头). 覆盖最多事例的最广假设为 $\{P(X_1, \dots, X_n)\}$, 其中 X_i 互不相同.

BDI 算法的基本思想是: (1) 从事例集中随机选取若干正例, 在这些事例中, 施用普化规则, 产生排斥所有反例的假设子句并以 AC 优选; (2) 施用特化规则于最广假设, 产生排斥所有反例的假设子句并以 WAC 优选; (3) 从产生的假设子句中再优选出最好的假设, 并使假设的优先值等于它所覆盖的正例数; (4) 去掉已被假设所覆盖的正例以及冗余假设. 重复上述过程, 直至正例空. 算法 BDI 是由自上而下和自下而上两种基本 ILP 综合而成的学习算法. 与传统 ILP 算法相比, BDI 具有较高的学习效率. 特别地, BDI 将归纳产生的子句的优先值赋予各自所覆盖的正例数, 使得子句的优先值在应用中具有明确的意义.

3 专家系统应用实例

1996 年, 国家 863 高科技项目计划在东北地区开展农业信息技术示范工程, 哈尔滨工业大学与绥化市承担大豆高产综合专家系统的研制任务. 该系统的核心工作之一是不完全信息推理与知识获取. 本节介绍本文的理论与算法的应用情况.

3.1 不完全信息推理应用

系统中各知识由两种途径得到: (1) 利用 BDI 算法从当地农业实验中进行归纳学习得到; (2) 直接由专家提供, 随问提供各优先值. 这类知识因人而异, 反映了专家的实际感受, 其主要原因是, 不同地点的气候与土壤条件差异极大. 不同专家对同一问题的不同感受反映了问题的各个侧面, 对知识逐一进行甄别是不现实的. 本系统将此类知识一并收入, 知识库也因此难以保证一致性, 但系统知识是丰富多样的. 采用通常的不确定推理技术都会

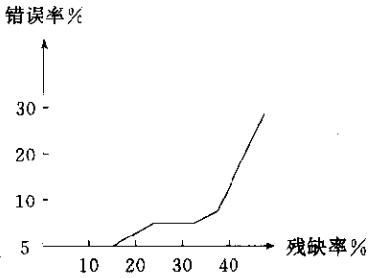


图1 优先解释推理实验结果

产生诸如知识完全性一致性要求与严格的信度限制等困难. 本文的优先逻辑程序上的优先解释方法却避免了上述困难. 本文的方法对各种可能的结论(解释), 利用已有知识对其进行综合评判, 进而产生优结论(解释).

为反映优先解释的准确性, 我们利用大豆病虫害知识与事实进行了多组残缺事实的推理实验, 结果如图 1 所示. 证据缺少 15%, 几乎不影响正确结论的获得; 证据缺少 40% 以上时, 结论错误率显著上升. 其他情况错误率维持在较低水平. 因此, 优先解释推理方法具有实用性.

3.2 知识获取

BDI 算法除用作直接从示例中学习以补充专家知识不足外, 另一作用是对已有知识进行渐进学习. 具体做法是将各类知识的训练示例加以保存, 在系统正常使用, 出现以下问题时, 调用 BDI 进行知识修正.

(1) 出现新的分类. 即当前事实导致现有知识中尚无的新类. 将当前事实做成新类的正示例, 并形成优先值为 1 的简单子句加入知识库中.

(2) 知识不足. 即当前事实出现后, 优先解释不能作出期望的结果. 若期望当前事实归属第 i 类, 则将当前事实做成新的正例加入第 i 类示例中, 直接调用 BDI 进行第 i 类的学习, 新产生的子句可能导致原来第 i 类中的子句冗余, BDI 能够删除冗余子句.

(3) 结论不正确. 即当前事实出现后, 正确的结论没有得到最高优先值. 若期望当前事实归属第 i 类, 则将当前事实做成新的正例加入第 i 类示例中, 对不正确的分类知识调用 BDI 重新学习, 学习中其他类的示例为正在学习类的反例.

不断地使用 BDI 进行上述过程的学习, 使系统知识渐进地满足分类要求.

4 结 论

针对不完全信息推理这一长期困扰知识工程的难题, 本文提出了一个基于优先解释的不完全信息推理框架. 这一理论依据知识对其解释进行综合评优, 使优先解释中具有较大优先值的事实受到系统中重要知识的直接支持. 事实表明, 优先解释不仅在理论上是知识的较佳逼近, 在应用中也是专家直觉的直接刻画.

作为对优先逻辑程序自动获取的支持, 本文基于 ILP 学习理论提出了一种结合自顶向下和自下而上两种基本归纳方法的学习算法 BDI. 该算法在启发函数作用下具有较强的搜索能力, 提供了对知识自动获取及理论渐进修正的直接支持.

本文介绍了优先理论及算法 BDI 在农业专家系统中的实际应用情况. 实践表明, 本文提出的方法在推理与知识获取方面性能优良, 不失为这一应用领域的满意的解决方案.

参考文献

- 1 Reiter R. On closed world databases. In: Gallaire H, Minker J eds. Logic and Database. New York: Plenum Press, 1978
- 2 Reiter R. A logic for default reasoning. Artificial Intelligence, 1985, 13(1): 81~132
- 3 McCarthy J. Circumscription—a form of nonmonotonic reasoning. Artificial Intelligence, 1985, 13(1): 27~39
- 4 Pinkas G. Reasoning, nonmonotonicity and learning in connectionist networks that capture propositional knowledge. Artificial Intelligence, 1995, 77(2): 203~247
- 5 Muggleton S, DeRaedt L. Inductive logic programming: theory and method. Journal of Logic Programming, 1994, 19(20): 629~679
- 6 Muggleton S, Buntine W. Machine invention of first order predicates by inverting resolution. In: Muggleton S ed. Inductive Logic Programming. London: Academic Press. 1992. 161~280
- 7 Lavrac N, Dzeroski S, Bratko I. Handling imperfect data in inductive logic programming. In: DeRaedt L ed. Advances in

Inductive Logic Programming. Amsterdam: IOS Press, 1996. 48~64

Precedent Interpretation Based Incomplete Information Reasoning and Its Application

YE Feng XU Xiao-fei WANG Ya-dong

(Department of Computer Science and Engineering Harbin Institute of Technology Harbin 150001)

Abstract Approximate reasoning with the incomplete information is one of the difficulties that the knowledge engineering has faced. A precedent logic program theory with the property of nonmonotonicity is proposed in this paper. The synthesis evaluation for the interpretation of knowledge can be taken with the theory, such that the optimal selection of interpretation is made possible which becomes the best approach to the current knowledge. The theory completion in the significance of optimal selection is achieved and the requirement of completion and consistency of knowledge are avoided. To acquire the precedent logic programs in the applications, based on an inductive logic programming, learning algorithm is presented which incorporates the multiple inductive methods and has greater ability of induction. The presented theory and the algorithm have been applied in an expert system and gained satisfactory results.

Key words Expert system, incomplete information reasoning, precedent interpretation, inductive logic programming.