

知识概念的表达与处理*

范慧琳

(华侨大学计算机科学系, 泉州 362011)

摘要 讨论概念的基本特征, 并给出概念间相互约束的模型. 针对解释学习中概念的自然形成过程, 提出一种基于模糊集合论的概念描述方法.

关键词 概念, 解释, 模糊

分类号 TP 18

概念是反映对象的特征或本质(以及本质属性)的思维形式. 它来源于感觉、知觉和表象, 但是它的产生是认识过程中的质变, 已经抓住了事物的本质、全体和内部联系了. 概念被认为是思维的基础, 与判断和推理并列为思维三要素, 故对思维的研究不可能脱离对人类概念系统的描述. 而人工智能中对概念的研究是在吸收其它学科对概念研究成果的基础上如何在计算机中形式化地建立模拟人类的概念系统的知识表达体系. 知识的表达形式之一是以概念作基本构成单元的语句, 因此概念的表达是研制基于概念的知识表达系统的基础. 传统的语义网络, 通过附有标记的连接建立符号间的相互联系, 可以看成是对概念的内涵和外延进行描述的工具. 而框架表达, 同样是通过 Slot-Value 结构来刻画现实世界中的实体概念. 但是概念的表达不是仅仅对概念的内涵与外延进行形式化处理就够了, 在机器学习, 自然语言理解等研究领域, 需要大型知识库作为系统运行的基础, 如何处理概念间的相互关系也十分重要. 为此, 本文首先简要地讨论概念的基本特征, 在此基础上给出了一个概念间相互约束的模型. 其次, 针对机器学习方法之一——解释学习中概念的自然形成过程, 提出一种基于模糊集合论的概念描述方法.

1 概念的基本特征及概念约束模型

一个概念的基本意义由其内涵和外延两个方面所决定. 内涵就是概念所反映的对象特征或本质, 外延是概念反映的那一事物或那一类事物的总和^[1]. 在目前的知识表达方法中, 概念的内涵一般用特征表来表示. 概念的外延由于是对概念所对应事物的穷举, 所以在一般的知识表达方案中都是通过上下位概念及抽象(即实例)来定义.

概念的内涵与外延是互相制约的, 二者具有反比关系. 即: 一个概念的外延愈大, 则它的内涵愈少; 反之, 一个概念的外延愈小, 则它的内涵愈多. 概念间可以相互约束正是反映了概念内

* 本文 1994-12-27 收到

涵与外延之间的变化过程,它反映了世界上事物的复杂性、运动过程、状态的变化等.我们把作为约束的概念记为 A ,把接受约束的概念记为 B .以 B 为主体,可把概念间的相互约束划分为三种基本模式:

(1) 事物性约束模式. 设“塑料”和“杯子”是二个基本概念,则复合概念“塑料杯子”反映了“杯子”的制造材料的特征已经被约束到了“塑料”这一特殊材料上,形成事物性概念;

(2) 动作约束模式. 如“张三看电视”。“看”的动作发出者被约束为一个特定的人,且“看”的对象也被约束到特定的物件上,成为动作概念;

(3) 状态约束模式. 如“天气冷了”. 通过状态的描述使得“天气”的气温属性被约束到概念“冷”上,成为状态概念.

由上述可见,概念的相互约束过程,就是概念某些特征受限制的过程. 即用逐渐增加内涵的方法使外延较窄的概念过渡到外延较宽的概念. 更复杂的相互约束可以在上面三种模式的基础上组合而成. 这种概念间的约束遵循以下两个规则.

规则 1 如果 B 的某一概念特征 Q 无约束, C 是 Q 可能的约束之一,则概念约束成立. 这里,概念 C 是 Q 的下位概念或从属概念,如“黑郁金香”. 郁金香具有颜色特征,但尚无具体约束,而“黑”是颜色的一种(具有从属关系),则“黑郁金香”是成功的一种约束.

规则 2 如果 B 的某一概念特征 Q 已经约束到某一概念 B_s 或概念集合 S 上,而 C 或是 B_s 的下位概念和从属概念,或是 S 中某一概念的下位概念和从属概念,则概念约束成立. 如:“不锈钢餐具”、概念“餐具”的制造材料特征可以是 $S = \{\text{钢材, 塑料, 玻璃, } \dots\}$,“不锈钢餐具”使“餐具”的制造材料限制在“不锈钢”(为钢材的下位概念)上,则“不锈钢餐具”是成功的约束之一.

作为约束的概念 A 与接受约束的概念 B 之间可否产生约束,可归结如附表.

附表 概念间的约束

A	B		
	要求被约束	要求不能被约束	未提及
要求约束	约束成立	约束要求不一致	约束成立
要求不能约束	约束要求不一致	约束不成立	约束不成立
未提及	约束成立	约束不成立	未知

当概念约束要求不一致时,应该及时修正. 这一情况以及未知情况的出现,表示着一种通过知识获取达到知识增长的要求.

2 概念的模糊化描述

基于解释的机器学习(EBL)是近年来的一个热门课题^[2], Mitchell 将 EBL 总结如下:给定:领域理论 DT (规则和事实集);目标概念 TC ;训练实例 TE ;可操作性准则 OC .

找出:满足 OC 的关于 TC 的充分条件.

EBL 的一般过程是,使用 DT ,找出 TE 为什么是 TC 之实例的证明(即解释),然后,根据 OC 对证明进行推广,从而得到关于 TC 的一个充分条件. 这一充分条件既是 TE 的一般化又是 TC 的特例化. 可操作性与不完善域论是 EBL 的关键问题. 可操作性往往不能以“可与不可”这种简单二值逻辑来回答,应视为模糊概念. 当域论不完善时,不完善域论对完善域论的贴

近程度也可用模糊取值来衡量^[3]. 因此, 模糊理论在 EBL 中大有可为. 以下仅就 EBL 中概念的实际形成过程, 给出一种基于模糊集合论^[4]的概念描述方法.

从 EBL 的“解释”的角度看, 一个概念实际上是大量的具有共同解释的实例的概括. 而概念 C 可用概念谓词 $C(x)$ 来表示, 意即“对象(或个体、实例) x 属于概念 C ”. 设 $C(x)$ 对应的实例集为 I , 对 I 中任一实例 $I_k (k \leq N)$, 可由其解释谓词 $Q_{ij}(I_k) (i, j \leq N)$ 的析取范式来描述

$$C(I_k) \Leftrightarrow \bigvee_i (\bigwedge_j Q_{ij}(I_k)),$$

意即“实例 I_k 属于概念 C 当且仅当 I_k 具有解释 $\bigvee_i (\bigwedge_j Q_{ij}(I_k))$ ”. 不失一般性, 仅取一个合取项来讨论, 即令

$$C(I_k) = \bigvee_i Q_{ij}(I_k),$$

其中 $Q_i(I_k)$ 为 I_k 的第 i 个解释谓词.

首先我们引入实例 I_k 的解释集. 把集合 $E_{I_k} = \{Q_1(I_k), Q_2(I_k), \dots, Q_i(I_k), \dots\}$ 称为实例 I_k 的解释集. 其中, 谓词 $Q_i(I_k)$ 仅做为集合的元素来讨论, 故可简记为 $E_{I_k} = \{Q_1, Q_2, \dots, Q_i, \dots\}$.

一般地, 观察概念 C 的大量实例 $I_1, I_2, \dots, I_k, \dots$ 所对应的解释集往往不尽相同, 因此, 如何在每一 E_{I_k} 的基础上较准确地取得 C 的解释集是个带模糊性的问题^[5]. 为此, 我们把解释谓词 Q 用来解释概念 C 的概率 $P_C(Q)$ 定义为 Q 出现在 $\bigcup_{I_k \in I} E_{I_k}$ 中的概率.

进一步地, 设 W 为一阶谓词全集, 定义概念 C 的模糊解释集 E_C 为 W 上的一个模糊子集, 其隶属函数为

$$\mu_{E_C}: W \rightarrow [0, 1], \mu_{E_C}(Q) = P_C(Q),$$

而 E_C 的截集 C_λ 即可定义为

$$C_\lambda = \{Q | P_C(Q) > \lambda, 0 \leq \lambda \leq 1\}.$$

至此, 概念 C 便可由其模糊解释集 E_C 或其截集 C_λ 中的谓词来描述. 某个体 x 在 λ 意义下属于概念 C , 即可定义为 $C(x) = \bigwedge_{Q(x) \in C_\lambda} Q_i(x)$. 概念的形成过程及概念的模糊解释集的形成过程如图 1 所示.

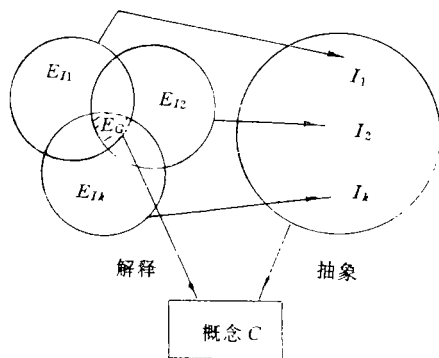
此外, 一个概念本身可能又是另一个概念的解释谓词, 同时, 一个解释谓词又有自己的解释谓词. 这样, 可用以解释谓词作为结点的树型结构——模糊解释树 T 来描述一个概念. 树 T 可递归定义如下:

(1) 一个事实是真值为 1 的单结点解释树;

(2) 解释树可由一个根(概念谓词)及若干个可取模糊值的与根相连的子概念解释树构成. 这些子树的根(子概念谓词)构成概念谓词的模糊解释集;

(3) 一棵概念解释树由有限个子概念解释树构成.

概念 $C-C(x)$ 取值的模糊性便表现在其解释集是一模糊集合, 即概念及其对象的解释是在



附图 概念及概念模糊解释集的形成过程

$[0,1]$ 上取值. 当我们欲求对象 x 属于概念 C 的真值 $\nu((C(x)))$ 时, 可以考虑为: 仅当概念 C 的模糊解释集 E_C 中 $\mu_{E_C}(Q)=1$ 这样的必要谓词 Q 的真值 $\nu((Q(x)))=0$ 时, $\nu((C(x)))$ 的值才为 0, 其它情况则按加权平均办法求 $C(x)$ 的真值, 即可定义如下

$$\nu((C(x))) = \begin{cases} 0 & \text{当 } Q \in C_\lambda \wedge \mu_{E_C}(Q) = 1 \wedge \nu(Q(x)) = 0, \\ \left[\sum_{Q \in C_\lambda} \nu(Q(x)) * \mu_{E_C}(Q) / \sum_{Q \in C_\lambda} \mu_{E_C}(Q) \right] & \text{其它,} \end{cases}$$

当给定一对象的若干解释谓词并建立起相应的概念解释树 T , 则该对象属于某概念的真值即在 T 上自底向上求出的各结点直至根结点的真值.

3 结束语

概念是人类对世界认识所形成的一种基本模式, 是人类知识组成的最小单元, 以概念作基本构成单元的语句是知识的表达形式之一, 而语句中往往存在大量的概念相互约束关系, 概念约束模型为研制基于概念的知识表达系统提供了一定的基础. 而在知识系统中, 要严格地描述一个概念往往是不现实的, 采用基于模糊集合论的不确定知识表示是一种良好的途径. 文中给出了针对 EBL 的概念模糊描述方法, 从模糊角度引入了概念的模糊解释集进而建立概念的模糊解释树, 并对概念的模糊值作了简要探讨, 为形成具有模糊特点的知识库作了基础性的工作. 这些方面仍有许多进一步的研究问题尚待解决.

参 考 文 献

- 1 中国人民大学哲学系逻辑教研室编. 形式逻辑. 北京: 中国人民大学出版社, 1980. 20~45
- 2 石纯一. 解释学习进展. 软件学报, 1993, 4(5): 51~57
- 3 何新贵. 加权模糊逻辑及其广泛应用. 计算机学报, 1989, 12(6): 458~464
- 4 汪培庄. 模糊集合论及其应用. 上海: 上海科技出版社, 1986. 35~60
- 5 何新贵. 人工智能新进展. 北京: 清华大学出版社, 1994. 49~106

Expression and Processing of Concept in Knowledge

Fan Huilin

(Dept. of Computer Science, Huaqiao Univ., 362011, Quanzhou)

Abstract Beginning with a discussion on the basic features of concept in knowledge, a model of mutual binding between concepts is given. In connection with natural formulation of concept in explanation-based learning (EBL), a descriptive method based on fuzzy-set theory is proposed.

Keywords concept, explanation, fuzzy