

概念的渐进式认知理论与方法

李金海^{1,2)} 米允龙³⁾ 刘文奇^{1,2)}

¹⁾(昆明理工大学数据科学研究中心 昆明 650500)

²⁾(昆明理工大学理学院 昆明 650500)

³⁾(中国科学院大学计算机与控制学院 北京 100190)

摘 要 概念是知识表示的基本认知单元,它由外延和内涵两部分构成.由于概念的外延与内涵可以相互诱导,所以概念的外延和内涵中一旦有一个被确定下来,那么这个概念也就随之确定.概念认知是将属于这一概念的特征属性筛选出来,同时把不属于这一概念的特征属性排除,即通过确定内涵的方式获得概念,它采用特定的认知方法来完成概念的识别.当前,概念认知正逐渐借鉴认知科学领域中的一些研究思想,不断地完善自身理论与方法.然而,现有的概念认知方法要求假定概念认知算子具有完全认知功能,但现实中由于个体认知的局限性往往会出现认知不完全的现象,在此情况下要对概念进行全面认知则需漫长等待,这意味着在时限约束条件下可能无法及时获取足够的概念信息.在认知科学领域,一般采取问题分解求解的策略解决此类问题,即根据当前的局部有限信息做出相应的快速决策,为阶段性认知提供有效的及时参考,是渐进式思维的一种认知模式.受此启发,本文提出概念的渐进式认知理论与方法,以处理不完全认知条件下的概念获取问题.具体地,提出概念渐进式认知的认知机理,其功能主要包括概念信息的增删改操作,以模拟一般的认知过程;针对线索为对象集、属性集以及对象集和属性集的三种不同情况,分别给出概念渐进式认知算法,并对它们的时间复杂度和空间复杂度进行分析.特别地,为了实现近似概念认知,设计乐观悲观两种信息融合策略,使得对于任意对象集和属性集为线索的概念认知至少获得一个上界概念.最后,数值实验表明概念渐进式认知算法的有效性.

关键词 粒计算;概念认知;形式概念分析;概念格;认知算子

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2019.02233

Incremental Cognition of Concepts: Theories and Methods

LI Jin-Hai^{1,2)} MI Yun-Long³⁾ LIU Wen-Qi^{1,2)}

¹⁾(Data Science Research Center, Kunming University of Science and Technology, Kunming 650500)

²⁾(Faculty of Science, Kunming University of Science and Technology, Kunming 650500)

³⁾(School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100190)

Abstract Concepts are often considered to be the basic cognitive units of knowledge representation, and they consist of two parts; one is the extent and the other is the intent. Considering that the extent and the intent of a concept are able to induce each other, we can obtain a concept once its extent or intent is completely determined. As is well known, concept cognition is to select all the attributes that belong to the concept, and at the same time, eliminate other attributes that do not belong to the concept. In other words, a concept can be formed by means of specifying its intent. Note that concept cognition achieves the task of recognizing the extent and the intent of a concept based on certain cognitive methods. In recent years, concept cognition has gradually been integrating the novel research ideas from the field of cognitive science in order to enrich the theory and improve the method continuously. Although the existing concept cognition methods are able to

meet different practical requirements, their concept-forming operators can only be effective under the special circumstance that they must satisfy the function of complete cognition. In the real world, however, the phenomenon of incomplete cognition of concept-forming operators frequently appears due to individual cognitive limitations. In this case, a comprehensive recognition of concepts often means a long time of waiting or delay. As a result, it may be impossible to acquire sufficient information for understanding the current concept well within the required time. In the field of cognitive science, decomposition and solving strategy was often adopted to deal with this kind of problems. In other words, it can provide timely reference for the current phased recognition by taking quick action based on the available local finite information, which belongs to the cognitive mode of incremental thinking. Motivated by this issue, this paper puts forward theories and methods for incremental cognition of concepts so as to solve the problem of acquiring concepts from incomplete cognitive environment. More specifically, we first propose cognitive mechanism for incremental concept cognition whose functions mainly include addition, deletion and revision operations for updating the current concept information. And then, using these useful operations, we can simulate general cognitive processes. What is more, we develop effective algorithms to implement incremental concept cognition under the circumstance where the given clue is in the form of an object set, an attribute set, or a pair of object and attribute sets. In addition, we also analyze the time complexity and the space complexity of these incremental cognition algorithms. Especially, when exact cognitive concepts cannot be obtained, we design two novel strategies (i. e. optimistic and pessimistic information fusion) to search for approximate cognitive concepts. This makes us at least find an upper concept for any given clue in the form of a pair of object and attribute sets once we are not able to obtain a lower concept in some particular cases. Finally, we conduct some numerical experiments to demonstrate the effectiveness of the proposed incremental cognition algorithms.

Keywords granular computing; concept cognition; formal concept analysis; concept lattice; cognitive operators

1 引 言

概念是知识表示的基本认知单元,它常被用来识别现实世界中的具体实体,或感知世界中的抽象对象^[1]. 研究概念可从不同的角度或侧面进行. 如从代数角度研究概念的层次结构^[2],从算法复杂性角度研究概念格构造^[3-5],从机器学习角度研究概念学习^[6],从认知心理学角度研究概念认知^[7],从粒计算角度研究概念粒化及粒度转移方式^[8-9],从云模型角度计算概念的外延与内涵^[10]. 本文主要关注概念认知,具体围绕概念的渐进式认知问题展开探讨.

概念认知是将属于这一概念的特征属性筛选出来,同时把不属于这一概念的特征属性排除,它采用特定的认知方法完成概念外延与内涵的识别. 目前,概念认知已成为一个新兴的研究方向. 比如,文献^[11]提出研究概念认知的三个具体层面:抽象层

面、机器层面和大脑层面,它们既是相对独立又是相互依赖促进的. 又如,文献^[12]在模糊环境下利用充分必要信息粒讨论了概念的不确定性认知,且仿真实验表明不确定性认知获取概念的效果较好. 再如,文献^[13]通过概念粒计算系统建立概念的迭代认知数学模型,它能够对各种类型的概念(经典的形式概念、推广的模糊概念等)进行认知. 实际上,粒计算在概念认知学习过程中起到了粒化对象(样本)或属性(特征)的作用,这是粒计算理论处理大数据的核心思想^[14-15]. 需要指出的是,在很多情况下,人们很难完成精确的概念认知. 例如,当提供的线索本身就存在不一致性时(如对象与属性之间相对于形成概念的外延与内涵匹配性较差^[16]),可以采取“退而求其次”策略以获得概念认知的近似结果,即在求解问题的复杂性与计算精度之间做出权衡. 此外,文献^[17]通过上、下近似逼近思想给出了概念认知的近似范围(即下确界概念和上确界概念组成的区间),进一

步讨论概念认知的精度,从而开启了近似概念认知研究.另外,对于快速增长的大数据,因时限性、多样性、复杂性等因素^[18],当前收集并存储的数据往往存在不完备、不确定等特点.此时进行精确的概念认知相当困难,但又不得不作出决策,因为在一定时间内得到尽可能满意的解,并根据满意解来及时指导下一步的行动通常是必要且紧迫的^[19].值得一提的是,从粒计算角度而言,根据当前不完备或不确定的信息作出及时决策以获得满意解,从而指导下一步的行动,并随着信息的进一步完备化或不确定性的逐渐降低,对已获得满意解作出更新以逼近最优解,这类问题可以通过由粗粒度到细粒度(coarse-to-fine)的策略来完成建模.当然,这种粒度变化策略还需结合具体的问题背景进行考虑,具有较大的灵活性,有兴趣的读者可查阅文献^[20-21].实际上,由粗粒度到细粒度是多粒度认知的一种典型代表模式,也是解决渐进式优化与认知问题的核心思想,详细的阐述见文献^[22-27].

然而,现有的概念认知研究工作^[7,12-13,16-17,28-29]在实施概念认知时,均假定认知算子具有完全认知功能,即由对象到属性(或由属性到对象)的认知过程能够一次性直接完成,它仅与给定数据或信息本身有关,不考虑时间、空间、环境等诸多因素对认知过程产生的影响.然而,现实世界中,由于个体偏好、先验知识、背景知识、时间、空间、环境等众多因素的影响,概念认知往往无法保证一次性就能够完成.比如,以对象识别共性属性为例,当从一批给定对象中找出它们共同拥有的所有属性时,一次性就直接找出所有的共性属性往往比较困难.实际上,从认知心理学角度来讲,找到事物所有共性属性(即认识清楚事物)是一个非常漫长的过程,因为个体通常受到自身认知的局限性而导致识别共性属性将是一个不断完善的反复过程^[30].这与现实中完全识别清楚一个概念的外延与内涵需要经历一个漫长的阶段也是一致的.为了描述方便,本文将这种不能一次性从给定对象中找出所有共性属性的问题称为不完全认知条件下的概念获取.

受上述讨论启发,本文尝试提出不完全认知条件下的概念渐进式认知理论与方法.在具体建模过程中,引入共性属性认知优先关系以反映个体偏好、先验知识、背景知识、时间等因素对概念认知造成的具体影响.在此基础上,针对线索为对象集、属性集以及对象集和属性集的三种不同情况,分别给出概念渐进式认知算法.在此基础上,讨论了基于共性属

性认知优先关系的认知算子与现有认知算子之间的差异,且数值实验表明本文的概念渐进式认知算法是有效的.

2 概念渐进式认知的基本原理

本文记对象集(即论域) U 的幂集为 2^U ,属性集 A 的幂集为 2^A .

经典的概念认知算子是一对集值映射^[31]:

$$f: 2^U \rightarrow 2^A,$$

$$g: 2^A \rightarrow 2^U.$$

通常,称 f 是对象-属性算子, g 是属性-对象算子.

定义 1. 对于属性-对象算子 g ,对象集 $X \subseteq U$,属性 $a \in A$,若 $X \subseteq g(a)$,则称 a 是 X 的一个共性属性.

那么,对象-属性算子的像 $f(X)$ 可看作 X 的所有共性属性组成的集合,属性-对象算子的像 $g(B)$ 则表示满足条件 B 的所有对象组成的集合.

在此基础上,易证 f 和 g 具有下列性质:

- (1) 若 $X_1 \subseteq X_2 \subseteq U$,则 $f(X_2) \subseteq f(X_1)$;
- (2) 若 $B_1 \subseteq B_2 \subseteq A$,则 $g(B_2) \subseteq g(B_1)$;
- (3) $X \subseteq g(f(X))$, $B \subseteq f(g(B))$.

从认知的角度,上述三个性质均有明确的语义解释:性质(1)表示,对象(样本)越多,其共性属性(特征)越少;性质(2)表示,共性属性越多,满足它们的对象越少;性质(3)表示,一个对象集经过复合认知算子 gf 作用后,对对象的认知比原来更充分;类似地,一个属性集经过复合认知算子 fg 作用后,对属性的认知也比原来更充分.

进一步,如果这两个认知算子满足 $f(X) = B$ 且 $g(B) = X$,那么称序对 (X, B) 为概念, X 为概念的外延, B 为概念的内涵.若一个概念 (X_1, B_1) 的外延包含于另一概念 (X_2, B_2) 的外延,则称 (X_1, B_1) 是 (X_2, B_2) 的子概念,或 (X_2, B_2) 是 (X_1, B_1) 的父概念.

此外,可以证明,序对 $(g(f(X)), f(X))$ 和 $(g(B), f(g(B)))$ 都是概念,即从一个对象集 X 或属性集 B 出发,经过两次迭代认知即可获得一个概念.从数学上看,这是相当漂亮的结果,这里的复合认知算子 gf 和 fg 分别相当于 U 和 A 上的闭包算子.但是,从认知心理学、机器学习、大数据等角度而言,这些复合算子的具体实施是很困难的,因为受时限性、多样性、复杂性等因素影响,对概念信息的理解和获取都是一个漫长的过程^[18,30].下面从认知心

理学角度给出一个解释. 在认知心理学中, 对认知算子 f 的理解比较复杂(其复合算子就更加复杂), 它与认知时刻及认知持续时间长短等因素均有关, 即涉及认知过程的各个方面. 比如, 假设 X 是由两幅图片组成的对象集, 那么早上查找两幅图片的共同特征 5 min 与下午查找两幅图片的共同特征 5 min, 它们的效果(即 $f(X)$)是不一样的; 另外, 晚上查找两幅图片的共同特征 5 min 与晚上查找两幅图片的共同特征 10 min, 其效果也是不同的. 换言之, 这类问题仅用 $f(X)$ 是无法清楚描述具体细节的.

鉴于此, 有必要将时间等因素反映到认知算子 f 上, 即需要对其作一些改进. 本文用 $f_t(X)$ 表示认知算子 f 在 t 时刻对对象集 X 认知一段时间后得到的结果, 即通过引入变量 t 来反映概念信息获取是一个漫长的过程. 为了区分于 f , 称 f_t 为渐进式认知算子.

需要指出的是, 对于属性-对象算子 g , 本文的理解与经典的情况保持一致, 原因是给定具体条件 B (属性或特征)筛选出特定对象 $g(B)$ 的问题, 通常认为这样的操作是可以快速实现的. 它与已知对象集 X 识别其“共性属性”是不一样的, 因为“共性属性”通常不容易一次被认知完整^[30].

定义 2. 给定渐进式认知算子 f_t , 对象集 $X \subseteq U$, 若对于任意 $\Delta t > 0$, 有 $f_{t+\Delta t}(X) = f_t(X)$, 则称 $f_t(X)$ 在 t 时刻对对象集 X 实现了完全认知; 否则, 称 $f_t(X)$ 在 t 时刻对对象集 X 的认知是不完全的.

那么, 从完全认知与否的角度而言, 经典形式概念分析理论^[31], 可以认为是基于完全认知建立的. 具体地, 给定对象集 X , 经过对象-属性算子 f , 就能够得到 X 的全体对象共同拥有的所有属性; 同理, 给定属性集 B , 经过属性-对象算子 g , 也能够得到拥有 B 中全体属性的所有对象. 换言之, 有了 X, B, f 和 g , 就可以直接假定 $f(X)$ 与 $g(B)$ 均是已知的. 实际上, 这种研究方式, 是通过构造性方法实现的, 当然得到的结果肯定也是完全认知概念.

注意, 本文仅将变量 t 引入概念认知中, 以表明概念认知算子受某些因素的影响而出现不完全认知现象. 这种处理方法仅仅只是为了区分于完全认知. 为了叙述方便, 本文称 f 为完全认知算子, f_t 为渐进式认知算子.

实际上, 渐进式认知算子 f_t 与完全认知算子 f 是对概念认知过程的不同解读所导致的. 具体地, f 强调完全认知, 在这种情况下, 获取概念只需经过两次完全认知: 第一次完全认知为 $X \rightarrow f(X)$, 第二

次完全认知是在第一次完全认知的基础上进行的, 即 $f(X) \rightarrow g(f(X))$, 进而获得完全认知概念 $(g(f(X)), f(X))$. 然而, f_t 则与时间等因素有关, 它强调不完全认知或渐进式认知, 并且到底需要经过多少次渐进式认知才能获得精确的概念外延与内涵事先是不知道的, 它在渐进式认知迭代过程中需要结合具体的问题背景来判断.

综上可知, 所谓概念渐进式认知, 是指通过增量的方式逐渐识别清楚一个概念. 它是一个非常复杂的过程, 持续时间事先是未知的, 且涉及认知过程的增删改操作, 而不是简单的单调递增过程. 基于上述理解, 下面讨论认知过程的增删改操作的设计问题.

首先, 考虑增量操作, 即后一次认知 $f_{t+1}(X)$ 是在前一次认识 $f_t(X)$ 的基础上通过新增认知部分 $F_{t+1}(X)$ 来完成的. 具体地,

$$f_{t+1}(X) = f_t(X) \cup F_{t+1}(X) \quad (1)$$

其中 $F_{t+1}(X)$ 表示 $t+1$ 时刻对对象集 X 再次识别后新增的认知部分. 为了严谨起见, 当 $t=0$ 时, 令 $f_t(X) = \emptyset$, 即初始状态为 $f_1(X) = F_1(X)$. 从式(1)不难发现, 已被识别的部分 $f_t(X)$ 不需要重新再认知, 新增的认知 $F_{t+1}(X)$ 被认为来源于对剩余属性特征的进一步识别.

注意, 执行式(1)中的增量操作, 它是有前提条件的, 否则概念渐进式认知就退化成一个简单的单调递增过程, 这与实际情况不符. 比如, 当式(1)中的 $F_{t+1}(X)$ 为空时, 即 $t+1$ 时刻对剩余属性特征的进一步识别没有发现新知识, 这个渐进式认知过程就无法继续下去. 本文将这个特殊情形视作需要修正概念认知的时刻.

具体地, 设计修正操作如下: 当新增认知部分 $F_{t+1}(X)$ 为空时, 对上一次的累积认知结果 $f_t(X)$ 进行修正:

$$f_t(X) \leftarrow f_t(X) - F_t(X) \quad (2)$$

同时, 把 $t+1$ 时刻的认知结果 $f_{t+1}(X)$ 退回到修正之前的水平, 即

$$f_{t+1}(X) \leftarrow f_{t-1}(X) \quad (3)$$

为了严谨起见, 约定修正操作至少在已执行过一次增量操作之后才允许实施, 即 $t \geq 1$.

综上, 当新增认知部分 $F_{t+1}(X)$ 不为空时, 执行增量认知操作; 否则, 执行修正认知操作.

需要指出的是, 修正操作实际上包含两个子操作: 式(2)可看作是删除操作, 即删除上一次新增的认知部分; 式(3)可看作是修改操作, 即将最新累积认知修改成 $t-1$ 时刻的认知结果. 再结合增量操

作,那么渐进式认知过程就具备对概念信息进行增删改的功能.

此外,考虑到修正操作会将之前认知过的部分又删减掉,所以 $F_{t+1}(X)$ 允许与之前的新增认知取相同值(一旦取重复值,意味着这两个状态之间的某个时刻已执行过修正操作,如 $F_t(X)=F_{t-2}(X)$,则表明 $t-1$ 时刻执行了修正操作).这与现实中概念渐进式认知存在反复现象是一致的,即删减掉的部分随着认识越来越深刻又会被重新添加回来.

由于增量操作与修正操作可以交替进行,那么如果没有其它约束,它们将一直持续下去(最坏的情况就是形成交替死循环),这显然不符合实际情况.为此,本文引入额外条件 $F_t(X)=F_{t-2}(X)$ 作为概念认知的暂停条件,以获得阶段性的概念认知结果.

定义 3. 设 g 是属性-对象算子, F_t 是新增认知算子, f_t 是渐进式认知算子.对于 $X \subseteq U, B \subseteq A$,若存在正整数 $t \geq 3$ 使得

$$F_t(X)=F_{t-2}(X) \tag{4}$$

$$f_t(X)=B, g(B)=X \tag{5}$$

则称序对 (X, B) 为渐进式认知概念.

上述定义中,式(5)比较容易理解,它与完全认知算子的概念定义是类似的;式(4)表示 $t-1$ 时刻实施过一次修正操作,但相对于未修正之前的状况,概念认知未取得任何进展,即此次修正无效.本文将该时刻视作已经获得阶段性的概念认知结果.

尽管如此,渐进式认知概念,随着时间的推移,后续再度实施修正操作,仍有可能获得更多的知识,即多次修正奏效.

此外,不管是增量操作,还是修正操作,渐进式认知算子的通项 $f_{t+1}(X)$ 都可以由新增认知算子的通项 $F_{t+1}(X)$ 迭代之后完全确定.因此,渐进式概念认知的核心是如何确定每个时刻的新增认知 $F_{t+1}(X)$.

为了解决这个问题,下面先从算子 $f_t(X)$ 和 $f(X)$ 的角度,讨论渐进式认知概念与完全认知概念的联系与区别.

(1) 当 t 足够大时,累积渐进式认知 $f_t(X)$ 与完全认知 $f(X)$ 可认为是等同的,即承认概念认知的有限性,最终趋同.

(2) 现实世界中,人们一般等不到 t 足够大时才进行概念认知,很多时候概念认知具有时效性,即获得一部分信息便开始实施概念认知,所以概念的渐进式认知与时间等因素密切相关,而概念的完全认知则不考虑这类干扰因素.

(3) 渐进式认知概念 (X, B) 与完全认知概念不一定相等.实际上,根据定义 3,渐进式认知概念 (X, B) 仅仅只是满足实施一次修正操作之后,未能获得更多的知识,但据此并不能得出 $f_t(X)$ 和 $f(X)$ 就一定相等,因为实施一次修正操作未能奏效,不代表实施多次修正操作也无法奏效.

(4) 渐进式认知概念 (X, B) 只是一个相对概念,从某种程度上而言它是不稳定的,即随着时间推移,它有可能再次被更新.然而,完全认知概念是已有信息诱导出的稳定概念,除非信息出现更新,否则它被认为是最终的认知结果.

(5) 渐进式认知概念 (X, B) 是根据当前信息作出的一个阶段性概念认知结果,它很可能只是一个满意解,主要是为了通过该满意解来及时指导下一步的行动.

例 1. 表 1 给出了一个形式背景 $K=(U, A, D)$,其中对象集 $U=\{x_1, x_2, x_3, x_4, x_5\}$,属性集 $A=\{a, b, c, d, e, f\}$,表中数字“1”表示所在行的对象拥有所在列的属性,数字“0”表示所在行的对象不拥有所在列的属性.取 $X_0=\{x_1, x_2\}$,下面计算线索 X_0 的渐进式认知概念.

表 1 形式背景 $K=(U, A, D)$

U	a	b	c	d	e	f
x_1	1	1	1	1	1	0
x_2	1	1	1	1	1	0
x_3	0	1	1	0	0	0
x_4	0	0	1	1	1	0
x_5	0	0	0	0	1	1

假设 $F_t(X_i)$ 为 t 时刻随机选择 X_i 的一些共性属性组成的集合,这里 X_i 是通过属性-对象算子 g 作用于累积共性属性集 $f_{t-1}(X_{i-1})$ 上产生的.那么由式(1)、式(2)、式(3),概念渐进式认知的具体实施步骤如下:

① 令 $F_1(X_0)=\{c, d\}$,执行增量操作,即

$$f_1(X_0)=F_1(X_0)=\{c, d\},$$

$$X_1=g(\{c, d\})=\{x_1, x_2, x_4\}.$$

② 由于 $F_2(X_1)=\{e\}$,执行增量操作,即 $f_2(X_1)=f_1(X_0) \cup F_2(X_1)=\{c, d, e\}$ (此处不妨认为 $f_1(X_1)=f_1(X_0)$,因为 X_0 的共性属性 c 和 d 也是 X_1 的共性属性), $g(\{c, d, e\})=\{x_1, x_2, x_4\}=X_1$.

③ 由于 $F_3(X_1)=\varnothing$,执行修正操作,即

$$f_2(X_1) \leftarrow f_2(X_1) - F_2(X_1)=\{c, d\},$$

$$f_3(X_1) \leftarrow f_1(X_1)=\{c, d\}.$$

④ 由于 $F_4(X_1)=\{e\}$,执行增量操作,即

$$f_4(X_1)=f_3(X_1)\cup F_4(X_1)=\{c,d,e\},$$
$$g(\{c,d,e\})=\{x_1,x_2,x_4\}=X_1.$$

⑤注意到,此时已满足 $F_4(X_1)=F_2(X_1)=\{e\}$ 且 $f_4(X_1)=\{c,d,e\}$ 和 $g(\{c,d,e\})=X_1$ 均成立,所以 $(X_1,\{c,d,e\})$ 是一个渐进式认知概念.

上述计算中,修正操作会受新增认知算子 F_t 的影响.比如,本例步骤①中,如果令 $F_1(X_0)=\{c\}$,那么步骤②就要开始执行修正操作,因为添加共性属性 c 之后,其它属性就无法再添加进去,所以需要通过修正操作退回到初始状态,以便调整 $F_t(X_0)$ 的赋值,获得更加理想的概念认知结果.

下面讨论如何设计新增认知算子 F_t ,以揭示概念渐进式认知的主要研究内容.

这个问题大致可描述为:给定对象集 X ,从第 1 次到第 n 次认知,假设依次新增的认知部分为 $F_1(X),F_2(X),\cdots,F_n(X)$,那么它们之间有何关联?即具体取什么值?显然,当 $F_{t+1}(X)\neq\varnothing$ 时,说明它识别出 $F_t(X)$ 未曾识别的共性属性(已被修正过的认知除外).这是问题的关键所在,为什么有些共性属性在上一次被识别出来,而有些共性属性却要延迟到下一次认知中才能被识别呢?从认知心理学角度而言^[30],这是由于属性被识别存在不均衡的现象,即有些属性容易被识别,而另外一些属性相对难识别.这就说明共性属性被识别存在“认知优先关系”.换言之,当人们识别共性属性时,会出现一些共性属性优于其它共性属性被提前识别出现象.造成共性属性“认知优先关系”的原因很多,比如,先验知识、个体偏好、背景知识等.下面列举几个例子予以说明.

例 2. 以表 1 为例.假设恰好被目标对象集 X 所拥有的属性为 X 的理想共性属性,那么对于那些除了被 X 拥有之外还被其它对象所拥有的共性属性而言,“其它对象越多,意味着这个共性属性越不理想”就成为先验知识.取 $X=\{x_1,x_2\}$,则共性属性的渐进式认知过程为 $a\rightarrow b,d\rightarrow c,e$,因为 a 是理想共性属性(刚好与 X 完全匹配,第一次就被识别出来), b,d 是次理想共性属性(除了与 X 匹配,还被另外一个对象匹配,第二次才能被识别出来),而 c,e 是最不理想的共性属性(除了与 X 匹配,还同时被另外两个对象匹配,最后才被识别出来).这里的共性属性“认知优先关系”是由先验知识获得的.

例 3. 图 1 给出了两幅图片,每幅图片包含若干图形,每个图形内的英文字母表示该图形的填充颜色,其中 r:红色,b:蓝色,y:黄色,g:绿色,p:粉

色.现要求找出它们共同拥有的图形.显然,不同的人去完成这个任务,最终结果很可能是一样的,但是识别出所有相同图形的先后顺序往往会存在差异.比如,现在派甲、乙两位男同学去完成这个任务,他们对图形本身没有偏好,但是对颜色有偏好.假设甲对颜色的偏好依次是:粉色→蓝色→绿色,而乙对颜色的偏好依次是:蓝色→绿色→粉色.那么甲和乙识别出相同图形的先后顺序分别为:圆柱体→三角形→长方体,三角形→长方体→圆柱体.尽管最终的结果相同,但是认知过程是有差异的.这里的共性属性“认知优先关系”是由于个体对颜色的偏好导致的.

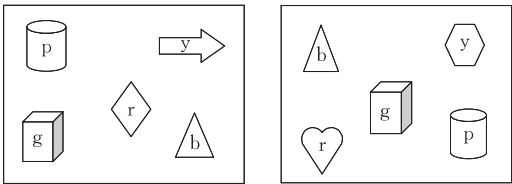


图 1 识别相同图形

例 4. 图 2 给出了两篇论文的参考文献,要求找出它们共同引用的论文.显然,这两篇论文的参考文献均是按“字母序”排列的.当查找相同引文时,可以充分利用这一背景知识.具体地,从上往下比对,第一篇找出来的相同引文是 Fehr 和 Gächter (2002),第二篇找出来的相同引文是 Liu (2016),第三篇找出来的相同引文是 Nowak (2006).这里的共性属性“认知优先关系”是由背景知识获得的.

论文1的参考文献:	论文2的参考文献:
[1] Apicella C L, Marlowe F W, Fowler J H, et al. Social networks and cooperation in hunter-gatherers. <i>Nature</i> , 2012, 481: 497-501	[1] Fehr E, Gächter S. Altruistic punishment in humans. <i>Nature</i> , 2002, 415: 137-140
[2] Fehr E, Gächter S. Altruistic punishment in humans. <i>Nature</i> , 2002, 415: 137-140	[2] Liu W Q. Public data evolution games on complex networks and data quality control. <i>Scientia Sinica Informationis</i> , 2016, 46: 1569-1590
[3] Liu W Q. Public data evolution games on complex networks and data quality control. <i>Scientia Sinica Informationis</i> , 2016, 46: 1569-1590	[3] Nowak M A. Five rules for the evolution of cooperation. <i>Science</i> , 2006, 314: 1560-1563
[4] Nowak M A. Five rules for the evolution of cooperation. <i>Science</i> , 2006, 314: 1560-1563	[4] Smith J M, Price G R. The logic of animal conflict. <i>Nature</i> , 1973, 246: 15-18

图 2 两篇学术论文引用的参考文献

根据上述实例及其讨论,下面给出共性属性“认知优先关系”的定义.

定义 4. 给定属性-对象算子 g ,对象集 $X\subseteq U$ 及其共性属性 $a,a_1,a_2\in A$.如果对于 X 的渐进式认知, a_1 与 a_2 存在优先关系,不妨记 $a_1>a_2$ (或 $a_2<a_1$),那么称 a_1 比 a_2 更容易被识别;如果 $a_1>a_2$ 且 $a_1<a_2$,那么称 a_1 与 a_2 被识别的难易程度相同.进一步,若 a 比剩余所有共性属性都更容易被识别,则称它是 X 最容易被识别的共性属性.

根据前面的讨论,这里的共性属性认知优先关系 $>$ (或 $<$) 一般由概念渐进式认知问题中的先验知

识、个体偏好、背景知识等诸多因素决定。实际上,共性属性认知优先关系并不是一成不变的。比如,执行修正操作后,通常会将修正部分无条件降为最不容易被识别的共性属性。

为了方便讨论,记 $\text{Ber}_t(X)$ 为 t 时刻对象集 X 的所有最容易被识别的共性属性组成的集合。

例 5. 以表 1 为例,假定目标对象集为 $X = \{x_1, x_2\}$ 。那么根据例 2 对渐进式认知的讨论结果可知,

$\text{Ber}_1(X) = \{a\}, \text{Ber}_2(X) = \{b, d\}, \text{Ber}_3(X) = \{c, e\}$ 。

定义 5. 给定对象集 $X \subseteq U$, 定义新增认知算子 F_t 为

$$F_t(X) = \{a \in \text{Ber}_t(X) \mid a \notin f_{t-1}(X)\} \quad (6)$$

其中 $f_{t-1}(X)$ 与 $F_1(X), F_2(X), \dots, F_{t-1}(X)$ 的关系由式(1)、式(2)和式(3)确定。

由定义 5 可知, $F_t(X)$ 表示 t 时刻对象集 X 最容易被识别的所有共性属性组成的集合, 这些共性属性在 t 时刻之前未被识别或已被识别但又被修正删除。至此, 关于如何选取新增认知部分 $F_t(X), F_2(X), \dots, F_n(X)$, 已讨论完毕。

利用新增认知 $F_t(X)$, 再结合式(1)、式(2)和式(3)可知, 在渐进式认知结束之前, 渐进式累积认知算子 f_t 满足以下条件:

$$f_t(X) = \bigcup_{i \in T} F_i(X) \quad (7)$$

其中 $T = \{1, 2, \dots, t\} - \{s_1, s_1 + 1, s_2, s_2 + 1, \dots, s_k, s_k + 1\}$, $s_1 + 1, s_2 + 1, \dots, s_k + 1$ 表示渐进式认知算子在 t 时刻之前执行过修改操作(即式(3))的所有时刻, 而 $s_1, s_2, \dots, s_k$ 表示执行修改操作时同时又实施删除操作(即式(2))的所有时刻。换言之, $f_t(X)$ 表示从初始时刻到 t 时刻, 对象集 X 的累积认知(部分认知由于被执行修正操作, 可能出现反复, 即添加后又被删除, 甚至再添加, 但该认知过程的总体趋势是渐进累积的)。

例 6. 以表 1 为例, 假定目标对象集为 $X = \{x_1, x_2\}$, 并满足例 2 中的先验知识。

① 当 $t = 1$ 时, $\text{Ber}_1(X) = \{a\}, F_1(X) = \{a\}, f_1(X) = F_1(X) = \{a\}$ 。

② 当 $t = 2$ 时, $\text{Ber}_2(X) = \{b, d\}, F_2(X) = \{b, d\}, f_2(X) = F_1(X) \cup F_2(X) = \{a, b, d\}$ 。

③ 当 $t = 3$ 时, $\text{Ber}_3(X) = \{c, e\}, F_3(X) = \{c, e\}, f_3(X) = F_1(X) \cup F_2(X) \cup F_3(X) = \{a, b, c, d, e\}$ 。

④ 当 $t = 4$ 时, $\text{Ber}_4(X) = \emptyset, F_4(X) = \emptyset$, 故 $f_3(X)$ 执行删除操作, $f_4(X)$ 执行修改操作 $f_4(X) =$

$$F_1(X) \cup F_2(X) = \{a, b, d\}.$$

⑤ 当 $t = 5$ 时, $\text{Ber}_5(X) = \{c, e\}, F_5(X) = \{c, e\}, f_5(X) = F_1(X) \cup F_2(X) \cup F_5(X) = \{a, b, c, d, e\}$ 。此时, $F_5(X) = F_3(X)$ 已满足, 所以结束概念渐进式认知。

从例 6 的③、④和⑤可知, $f_t(X)$ 在渐进式认知过程中, 确实会发生反复现象。

性质 1. 给定对象集 $X \subseteq U$, 则对任意 $t, f_t(X) \subseteq f(X)$ 。

证明. 根据上述讨论, 显然成立。证毕。

性质 2. 给定对象集 $X \subseteq U$, 则对任意 $t, X \subseteq g(f_t(X))$ 。

证明. 由式(7)可知,

$$f_t(X) = \bigcup_{i \in T} F_i(X),$$

再根据属性-对象算子 g 以及性质 1 的结论, 有 $g(f(X)) \subseteq g(f_t(X))$ 。注意到 $X \subseteq g(f(X))$, 即可证得 $X \subseteq g(f_t(X))$ 。证毕。

需要指出的是, 如果将性质 2 中的对象集替换为属性集, 则结论不成立, 即给定属性集 $B \subseteq A$, 对任意 $t, B \subseteq f_t(g(B))$ 不成立, 具体见以下反例。

例 7. 以表 1 为例, 并假定目标属性集为 $B = \{a, b, d\}$ 。那么根据例 2 和例 6 中渐进式认知的讨论结果可知,

$$g(B) = X = \{x_1, x_2\},$$

$$f_1(g(B)) = f_1(X) = F_1(X) = \{a\} \subset B.$$

性质 3. 给定属性集 $B \subseteq A$, 则存在 t 使得 $B \subseteq f_t(g(B))$ 。

证明. 根据约定, 本文承认概念认知的有限性, 即当 t 足够大时, $f_t(X)$ 与 $f(X)$ 趋同。因此, 存在足够大的 t , 使得 $f_t(g(B)) = f(g(B))$ 。由于 $B \subseteq f(g(B))$, 所以也有 $B \subseteq f_t(g(B))$ 。证毕。

下列例子进一步解释性质 3 的结论。

例 8. 以表 1 为例, 并假定目标属性集为 $B = \{a, b, d\}$ 。那么由例 2 和例 6 中渐进式认知的讨论结果可得,

$$g(B) = X = \{x_1, x_2\},$$

$$f_2(g(B)) = f_2(X) = F_1(X) \cup F_2(X) = \{a, b, d\},$$

$$B \subseteq f_2(g(B)).$$

3 概念的渐进式认知方法

本节在概念渐进式认知原理的基础上, 继续讨论如何从一个已知线索(即对象集, 属性集, 或对象

集和属性集)通过渐进式认知获取相应的概念.

3.1 线索为对象集的概念渐进式认知

假定以对象集 X_0 为线索,通过 f_1 作用可得 X_0 的第一次共性属性认知 $f_1(X_0)$,不妨令 $B_0 = f_1(X_0)$ 表示这些共性属性;利用属性-对象算子 g 对当前获得的共性属性集 B_0 进行对象认知,获得更新后的对象集 $g(B_0)$,不妨令 $X_1 = g(B_0)$ 表示更新后的对象集;进一步,通过 f_1 又可以获得 X_1 的第一次共性属性认知,根据迭代累积认知思想,如果未实施修正操作,当前认知结果应等于旧认知加上新增认知,即 $B_1 = f_1(X_1) \cup B_0$ (如果发生修正操作,意味着 X_1 重新被赋值,那么以最新一次赋值为准);再利用属性-对象算子 g 对共性属性集 B_1 进行对象认知,又可获得更新后的对象集 $X_2 = g(B_1)$. 上述过程进行反复迭代,可得到一系列的认知对象集 $X_3, X_4, \dots, X_n$ 和一系列的共性属性集 $B_2, B_3, \dots, B_n$. 为了下文叙述方便,本文将上述过程称为迭代累积认知.

应注意到,上述迭代累积认知过程,与第 2 节介绍的概念渐进式认知原理似乎不太一致. 原因是第 2 节强调的是针对同一对象集 X 进行共性属性的渐进式认知,而这里的迭代累积认知得到的是一系列对象集 $X_1, X_2, \dots, X_n$. 换言之,对相同对象集反复识别是第 2 节概念渐进式认知原理的核心思想,一系列对象集的迭代累积认知除非它们相等,否则第 2 节提出的概念渐进式认知原理应用于此处效果会不佳,即式(7)

$$f_t(X) = \bigcup_{i \in T} F_i(X)$$

不能较好的发挥作用(它是渐进式认知思想的具体表现). 令人欣慰的是,上述迭代累积认知产生的一系列有效对象集 $X_1, X_2, \dots, X_n$ (发生修正操作时,无效的对象集不予记录),确实是相等的,下面把这个结论通过性质 4 给出.

性质 4. 给定对象集 $X_0 \subseteq U$, 如果通过算子 f_1 和 g 对 X_0 进行迭代累积认知得到一系列的对象集 $X_1, X_2, \dots, X_n$, 那么 $X_0 \subseteq X_1 = X_2 = \dots = X_n$.

证明. 根据迭代累积认知过程的设计,可得 $X_1 = g(B_0) = g(f_1(X_0))$,再结合性质 2 的结论: $X_0 \subseteq g(f_1(X_0))$ 可知 $X_0 \subseteq X_1$.

进一步,当 $t \geq 2$ 时,因为 $X_{t-1} = g(B_{t-2})$, $B_{t-1} = f_1(X_{t-1}) \cup B_{t-2}$, $X_t = g(B_{t-1})$, 所以 $B_{t-2} \subseteq B_{t-1}$, 故有 $g(B_{t-2}) \supseteq g(B_{t-1})$, 即 $X_t \subseteq X_{t-1}$.

另一方面,注意到 $X_{t-1} \subseteq g(f_1(X_{t-1}))$, 因此 $X_t = g(B_{t-1})$
 $= g(f_1(X_{t-1}) \cup B_{t-2})$
 $= g(f_1(X_{t-1})) \cap g(B_{t-2}) \supseteq X_{t-1} \cap X_{t-1} = X_{t-1}.$

综上两方面可得 $X_{t-1} = X_t (t \geq 2)$, 即 $X_1 = X_2 = \dots = X_n$ 成立. 证毕.

性质 4 表明,式(7)中的渐进式认知 $f_t(X)$ 是完全适合迭代累积认知的,也能够发挥作用,因为迭代累积认知过程产生的一系列对象集都是相等的(初始对象集 X_0 除外).

下面讨论给定线索为对象集 X_0 ,通过渐进式认知最终将获得一个什么样的概念(X, B_t)? 这个问题与第一次有效的共性属性认知 $B_0 = f_1(X_0)$ 密切相关,因为这直接决定了最终概念的外延 $X = g(B_0)$. 当然,第一次有效的共性属性认知很可能是经过反复修正后得到的. 实际上,这与现实中的概念认知是吻合的,因为日常中人们一开始的认知方向一旦确定下来(可能是多次修正之后的结果),就基本固化了最终概念认知结果的总体趋势,后续进一步的迭代累积认知只是去逐步完善这个假想概念而已. 换言之,线索为对象集的认知,外延会被提前确定下来,而内涵则需经历一个漫长的过程才能完全认识清楚.

算法 1 给出了以对象集 X_0 为线索的概念渐进式认知,易证最终输出的序对 (X, B_t) 是渐进式认知概念. 一方面,由步骤 3、步骤 5 和步骤 6 可知, $F_t(X) = F_{t-2}(X)$, $B_t = f_t(X)$. 另一方面,由性质 4 可得, $g(B_t) = g(f_t(X)) = X$. 再根据定义 3,即可证得 (X, B_t) 是渐进式认知概念.

算法 1. 对象集 X_0 的概念渐进式认知.

输入: 对象集 X_0 , 算子 g 和 F_t

输出: 渐进式认知概念 (X, B_t)

1. 令 $i = 1$;
2. $B_0 = F_i(X_0)$, $X = g(B_0)$;
3. 如果 $F_2(X) \neq \emptyset$, 那么令 $F_1(X) = B_0$ 并转步骤 4; 否则, 当 $i \geq 3$ 且 $F_i(X_0) = F_{i-2}(X_0)$, 或概念认知已遍历所有的特征属性, 令 $B_t = B_0$ 并转步骤 8, 其它情况则执行 $i \leftarrow i + 1$ 并返回步骤 2;
4. 令 $t = 1$, $B_t = F_1(X)$;
5. 如果 $t \geq 3$ 且 $F_t(X) = F_{t-2}(X)$, 则转步骤 8;
6. 如果 $F_t(X) = \emptyset$, 那么
 $B_t = f_t(X) = B_{t-1} - F_{t-1}(X)$,
否则, $B_t = f_t(X) = B_{t-1} \cup F_t(X)$;
7. $t \leftarrow t + 1$, 返回步骤 5;
8. 输出 (X, B_t) , 算法结束.

另外,通过分析不难发现,算法 1 的时间复杂度为 $O(|A|(|U| + |A|))$, 空间复杂度为 $O(t)$.

最后,通过一个实例说明算法 1 如何实施.

例 9. 以表 1 为例,并假定对象集为 $X_0 = \{x_1\}$. 根据例 2 中概念渐进式认知的讨论结果可知,第一

次共性属性识别结果为 $B_0 = f_1(X_0) = \{a\}$, 且认知概念的外延为 $X = g(B_0) = \{x_1, x_2\}$. 由于 $F_2(X) = \{b, d\} \neq \emptyset$, 之后进入共性属性迭代累积认知过程:

$$\begin{aligned} B_1 &= f_1(X) = F_1(X) = \{a\}, \\ F_2(X) &= \{b, d\}, \\ B_2 &= f_2(X) = F_1(X) \cup F_2(X) = \{a, b, d\}, \\ F_3(X) &= \{c, e\}, \\ B_3 &= f_3(X) = F_1(X) \cup F_2(X) \cup F_3(X) \\ &= \{a, b, c, d, e\}, \\ F_4(X) &= \emptyset, \\ B_4 &= f_4(X) = F_1(X) \cup F_2(X) = \{a, b, d\}, \\ F_5(X) &= \{c, e\}, \\ B_5 &= f_5(X) = F_1(X) \cup F_2(X) \cup F_5(X) \\ &= \{a, b, c, d, e\}. \end{aligned}$$

由于 $F_3(X) = F_5(X)$, 所以结束概念渐进式认知算法, 并输出最终结果 $(X, B_5) = (\{x_1, x_2\}, \{a, b, c, d, e\})$.

注意, 上述计算中, 如果 X_0 的第一次共性属性认知结果为 $B_0 = f_1(X_0) = \{c\}$ 或 $\{e\}$, 那么 $X = g(B_0) = \{x_1, x_2, x_3, x_4\}$ 或 $\{x_1, x_2, x_4, x_5\}$, 就出现 $F_2(X) = \emptyset$, 此时第一次共性属性认知就需要执行修正操作之后才能确定结果.

3.2 线索为属性集的概念渐进式认知

众所周知, 给定属性集 B_0 , 通过属性-对象算子 g , 能够得到相应的对象集 $X_0 = g(B_0)$. 换言之, 当线索为一个属性集时, 可以通过 g 将概念渐进式认知问题归结为线索是一个对象集的情形, 即此处针对线索为属性集的情况, 不用再单独提出一种新的概念渐进式认知方法.

另外, 线索为属性集 B_0 时, 情况会更加特殊, 因为此时的概念渐进式认知, 它不像线索为对象集的情形: 需要通过一次有效的共性属性认知之后才能确定概念外延的大致情况. 这里的概念外延直接就由 $X_0 = g(B_0)$ 确定了. 因此, B_0 所对应的对象集 X_0 在整个概念渐进式认知过程中, 均不发生改变.

根据上述讨论, 线索为给定属性集 B_0 的概念渐进式认知, 首先通过属性-对象算子 g , 将其转化为线索为对象集 $X_0 = g(B_0)$ 的概念渐进式认知问题. 然后, 采用第 3.1 节提出的算法 1, 就可以获得渐进式认知概念 (X_0, B_i) .

例 10. 以表 1 为例, 并假定线索为属性集 $B_0 = \{b\}$. 首先, 将其转化为线索是对象集 $X_0 = g(B_0) = \{x_1, x_2, x_3\}$ 的概念渐进式认知问题. 然后, 类似于例 9, 通过算法 1 即可获得渐进式认知概念 $(X_0, B_4) =$

$(\{x_1, x_2, x_3\}, \{b, c\})$.

需要指出的是, 线索为属性集的概念渐进式认知, 与线索为对象集的概念渐进式认知有所区别的主要原因是: 本文事先假定对象-属性算子 f 遵循不完全认知, 而属性-对象算子 g 则遵循完全认知. 这导致经典形式概念分析中事先假定对象与属性的地位平等将不再成立. 实际上这也是概念渐进式认知区别于传统概念认知的关键所在, 这种差异不但体现在认知机理上, 还反映在不同类型的线索的处理上. 传统概念认知将对象集和属性集等同对待^[7, 12-13, 16-17, 28-29], 而概念渐进式认知则重点关注线索为对象集的情况, 造成两者差异的原因是从认知心理学角度识别对象的共性属性需要经历一个漫长的过程^[30].

3.3 线索包含对象集与属性集的概念渐进式认知

假设给定线索为对象集 X_0 与属性集 B_0 , 那么相对于线索只有一个对象集或属性集的情况, 概念渐进式认知的难度将大大增加. 具体原因分析如下:

首先, 注意到, 这里给定的两部分信息 X_0 与 B_0 在概念渐进式认知过程中是相互影响的, 彼此不能独立于对方 (即只关注其中之一) 而进行概念渐进式认知. 换言之, 既不是简单的重复第 3.1 节线索为对象集的概念渐进式认知, 更不是纯粹的重复第 3.2 节线索为属性集的概念渐进式认知. 此处需要提出新的概念渐进式认知方法.

其次, 在概念渐进式认知过程中, 充分考虑两方面 (即对象集与属性集) 的信息是非常必要的, 即需要通过信息融合去解决这个问题. 但是, 信息融合不能盲目进行, 要分析双方提供的信息是否满足融合限制条件 (这一点与经典概念认知类似, 不同之处在于本文将利用可匹配性作为信息融合的限制条件, 而经典概念认知则是通过下、上近似约束范围作为信息融合的限制条件^[16-17]). 因此, 在概念渐进式认知之前, 有必要先判断双方提供的信息是否具有有一定的匹配度 (下文将详细讨论这个问题). 当然, 对象集与属性集提供的信息出现不匹配 (即信息融合无法进行) 也是允许的, 此时也能够进行概念渐进式认知, 只不过最终的认知结果相对于可匹配的情形会有所欠佳而已 (本小节末将单独讨论这个问题).

然后, 如果对象集 X_0 与属性集 B_0 可以进行信息融合, 那么还应注意到概念认知学习的方式是多样化的, 获得什么样的认知结果与拟采取的信息融合策略密切相关. 这与认知心理学相关原理是一致的, 认知结果与认知过程及认知手段均有关.

最后,如果给定对象集 X_0 与属性集 B_0 ,它们无法做到精确认知,那么还得给出概念认知的精度.实际应用中,这种情况发生的可能性极大,因为概念能够做到精确认知的情形毕竟只是少数,大量的概念认知都是在不确定性环境下完成的.

基于上述分析,下面给出相应的处理方法.为了便于讨论信息融合,本文把属性集 B_0 提供的信息也转化成对象集 $Y_0 = g(B_0)$ 的形式,那么给定对象集和属性集的概念渐进式认知就变成如何对两个对象集 X_0 与 Y_0 进行信息融合以及概念认知的问题.在此之前,先对对象集 X_0 与属性集 B_0 之间的匹配性做一些约束,以增强概念渐进式认知的有效性.

定义 6. 给定对象集 X_0 与属性集 B_0 ,令 $Y_0 = g(B_0)$. 若 $X_0 \cap Y_0 = \emptyset$,则称对象集 X_0 与属性集 B_0 之间是不匹配的.

本文重点讨论给定对象集 X_0 与属性集 B_0 之间存在匹配的情形,且匹配程度约定为

$$\mu(X_0, B_0) = \frac{|X_0 \cap Y_0|}{|X_0 \cup Y_0|} \quad (8)$$

对于对象集 X_0 与属性集 B_0 不匹配的情形,本节未将单独讨论其概念渐进式认知问题.

当对象集 X_0 与属性集 B_0 存在匹配关系时,那么 X_0 与 $Y_0 = g(B_0)$ 在信息融合过程中应保持对等地位(即权重相同),这在信息融合公式的设计上表现为对称性成立.本文采取“一主一辅”的两种信息融合方式:

(1) 对象集 X_0 与 Y_0 悲观取交 $X_0 \cap Y_0$,它是主要信息融合机制;

(2) 对象集 X_0 与 Y_0 乐观取并 $X_0 \cup Y_0$,它是辅助信息融合机制.

实际上,主要信息融合机制(悲观取交)与 X_0 和 B_0 是否存在匹配关系是密切相关的.如果 X_0 和 B_0 不匹配,那么 $X_0 \cap Y_0 = \emptyset$,即悲观取交无实际意义.此外,对于悲观取交 $X_0 \cap Y_0$ 与乐观取并 $X_0 \cup Y_0$,若直接按第 3.1 节中提出的算法 1 进行概念渐进式认知,则获得两个概念,但这两个概念不一定存在关联性.为此,下面讨论悲观取交与乐观取并得到的两个概念,可以作为概念认知的近似结果.具体地,证明悲观取交 $X_0 \cap Y_0$ 得到的认知概念可以看作近似概念认知的下界,乐观取并 $X_0 \cup Y_0$ 得到的认知概念可以看作近似概念认知的上界.从这个角度而言,辅助信息融合机制(乐观取并)只是为了获得一个上界概念,以协助主要信息融合机制(悲观取交)产生的下界概念,对目标概念进行逼近.

① 悲观取交 $X_0 \cap Y_0$ 的概念渐进式认知

此时可以参考第 3.1 节的讨论结果,考虑概念渐进式认知问题,但仍有一些细微的差别需要强调.具体地,由于 $Y_0 = g(B_0)$,所以 $X_0 \cap Y_0$ 必拥有 B_0 中的所有属性.因此,很自然地就会把已确认的共性属性集 B_0 作为初始对象集 $X_0 \cap Y_0$ 第一次共性属性认知的结果.

算法 2 给出了相应的概念渐进式认知过程,最终输出的认知概念为 (Y_0, B_s) . 此外,通过分析不难发现,算法 2 的时间复杂度为 $O(|A|^2)$,空间复杂度为 $O(s)$.

算法 2. 对象集 $X_0 \cap Y_0$ 的概念渐进式认知.

输入: 对象集 $X_0 \cap Y_0$, 算子 g 和 F_s .

输出: 渐进式认知概念 (Y_0, B_s)

1. 初始化 $B_0 = F_1(X_0 \cap Y_0)$, $Y_0 = g(B_0)$;
2. 令 $s = 1$, $B_s = F_1(Y_0)$;
3. 如果 $s \geq 3$ 且 $F_s(Y_0) = F_{s-2}(Y_0)$,则转步骤 6;
4. 如果 $F_s(Y_0) = \emptyset$,那么

$$B_s = f_s(Y_0) = B_{s-1} - F_{s-1}(Y_0),$$
 否则, $B_s = f_s(Y_0) = B_{s-1} \cup F_s(Y_0)$;
5. $s \leftarrow s + 1$,返回步骤 3;
6. 输出 (Y_0, B_s) ,算法结束.

② 乐观取并 $X_0 \cup Y_0$ 的概念渐进式认知

给定对象集 $X_0 \cup Y_0$,直接采用第 3.1 节中提出的算法 1 进行概念的渐进式认知,不妨设最终得到的认知概念为 (X, B_t) .

下面,讨论悲观取交 $X_0 \cap Y_0$ 得到的认知概念 (Y_0, B_s) 与乐观取并 $X_0 \cup Y_0$ 得到的认知概念 (X, B_t) 之间的关系.

根据性质 4,有 $Y_0 \subseteq X_0 \cup Y_0 \subseteq X$ 成立.因此, (Y_0, B_s) 是 (X, B_t) 的子概念,故可将前者看作概念认知的下界,而把后者看作概念认知的上界.换言之,线索为给定的对象集 X_0 与属性集 B_0 ,通过概念的渐进式认知,最终获得一个下界概念 (Y_0, B_s) 和一个上界概念 (X, B_t) ,它们共同组成概念渐进式认知的结果.除非它们相等,否则这只是一个近似认知范围,即目标概念被认为介于下界概念与上界概念之间.当下界概念和上界概念不相等时,更加精确的认知结果暂时无法提供,它有待于信息的进一步完善.

定义 7. 给定对象集 X_0 与属性集 B_0 ,令 $Y_0 = g(B_0)$. 假设悲观取交 $X_0 \cap Y_0$ 得到的认知概念为 (Y_0, B_s) ,乐观取并 $X_0 \cup Y_0$ 得到的认知概念为 (X, B_t) . 若 $Y_0 \neq X$,则称概念认知是不精确的,并定义其精度为

$$\gamma(X_0, B_0) = \frac{|Y_0 \cap X|}{|Y_0 \cup X|} \quad (9)$$

现实中, 给定对象集 X_0 与属性集 B_0 , 大部分情况下概念的渐进式认知结果都是不精确的; 换言之, 出现精确认知的可能性较小, 获得近似认知的可能性较大。

另外, 还需要指出的是, $\gamma(X_0, B_0)$ 一方面刻画了对象集 X_0 与属性集 B_0 提供的信息的匹配程度; 另一方面, 也反映了概念渐进式认知结果的不确定性程度。精度只是提供某种程度上的一个度量参考, 不宜作为一个绝对的衡量指标。

例 11. 以表 1 为例, 并假定线索为对象集 $X_0 = \{x_1, x_2, x_3\}$ 与属性集 $B_0 = \{a\}$, 那么 $Y_0 = g(B_0) = \{x_1, x_2\}$, $X_0 \cap Y_0 = \{x_1, x_2\} \neq \emptyset$, $X_0 \cup Y_0 = \{x_1, x_2, x_3\}$. 因此, 对象集 X_0 与属性集 B_0 提供的信息是匹配的, 可以通过对悲观取交与乐观取并进行概念的渐进式认知, 以获得相应的认知概念信息。

考虑悲观取交 $X_0 \cap Y_0 = \{x_1, x_2\}$ 的概念认知。根据算法 2, 可以获得认知概念 $(Y_0, B_5) = (\{x_1, x_2\}, \{a, b, c, d, e\})$. 进一步, 考虑乐观取并 $X_0 \cup Y_0 = \{x_1, x_2, x_3\}$ 的概念认知。具体地, 由算法 1, 乐观取并 $X_0 \cup Y_0 = \{x_1, x_2, x_3\}$ 得到另一认知概念 $(X, B_4) = (\{x_1, x_2, x_3\}, \{b, c\})$. 综上, 线索为对象集 X_0 与属性集 B_0 , 通过概念的渐进式认知, 得到下界概念 $(Y_0, B_5) = (\{x_1, x_2\}, \{a, b, c, d, e\})$ 和上界概念 $(X, B_4) = (\{x_1, x_2, x_3\}, \{b, c\})$, 且它们的精度为

$$\gamma(X_0, B_0) = \frac{|Y_0 \cap X|}{|Y_0 \cup X|} = \frac{|\{x_1, x_2\} \cap \{x_1, x_2, x_3\}|}{|\{x_1, x_2\} \cup \{x_1, x_2, x_3\}|} = \frac{2}{3}.$$

最后, 讨论对象集 X_0 与属性集 B_0 之间不存在匹配关系时 (即 $X_0 \cap Y_0 = \emptyset$, 其中 $Y_0 = g(B_0)$), 如何进行概念的渐进式认知。

这种情况下, 主要信息融合机制 (悲观取交) 已无实际意义。原因是悲观取交 $X_0 \cap Y_0$ 为空集, 如果对其进行概念渐进式认知, 那么认知结果将是一个空概念。既然主要信息融合机制 (悲观取交) 已不起作用, 那么辅助信息融合机制也就没有了存在的必要, 因为构造它的初衷只是为了辅助主要信息融合机制产生的下界概念以逼近目标概念。

尽管如此, 取并思想 $X_0 \cup Y_0$ 依然可以借鉴。不妨设对象集 $X_0 \cup Y_0$ 经过算法 1 进行渐进式认知之后, 得到的概念为 (X, B_i) . 换言之, 当对象集 X_0 与属性集 B_0 之间不存在匹配关系时, 虽然信息融合不

再适用, 但仍然可以获得一个认知概念 (X, B_i) . 实际上, 这个概念是给定线索为对象集 X_0 与属性集 B_0 获得的最大泛化概念, 它是最大泛化意义上的“上界概念”。此时, 概念认知的精度无法给出, 因为它只有单侧概念。

4 概念认知方法的对比分析

本节将本文中线索为对象集、属性集、对象集和属性集的概念渐进式认知方法, 与现有的概念认知方法作对比分析, 以说明它们之间的异同。

(1) 研究背景

本文讨论概念渐进式认知, 其问题背景来源于认识事物需要经历一个漫长的过程, 具体表现为对象到属性的认知是一个渐进过程; 现有的概念认知^[12, 28-29]主要针对完全认知环境, 即对象到属性的认知算子不受时间、空间、环境等诸多因素的影响, 它是一种理想的状态。

(2) 研究目的

本文提出的概念渐进式认知方法, 其研究目的是希望针对特定的不完全认知环境, 建立渐进累积认知函数, 获得有效的概念渐进式认知模式, 以实现满意的阶段性认知, 并逐渐逼近完全认知概念; 现有概念认知方法^[12, 16-17, 28-29]针对完全认知环境模拟概念认知, 其研究目的是找到最终的认知概念, 不考虑概念认知算子受认知环境的影响, 更无法提供阶段性认知。

(3) 研究方法

本文提出的概念渐进式认知方法, 是通过“对象到属性”与“属性到对象”的双向迭代累积认知完成的。特别地, 当线索为对象集和属性集时, 采取“悲观取交”与“乐观取并”一主一辅两种策略融合对象集和属性集两方面的信息; 现有概念认知方法通过粒概念的粗糙下、上近似算子^[16-17, 29]或充分必要信息粒合成^[12, 28]完成概念获取。

(4) 研究结果

本文提出的概念渐进式认知方法, 其认知结果与拟采取的新增认知函数 (即渐进式方式) 以及信息融合策略 (线索为对象集和属性集) 均密切相关; 现有概念认知方法与粒概念的近似逼近能力^[16-17, 29]或粒合成方式^[12, 28]有关。值得一提的是, 这些方法的概念认知结果有可能两两不相同。

(5) 适用范围

本文提出的概念渐进式认知方法, 适用于给定

线索为任意对象集、属性集或对象集和属性集的情形,而文献[12,28]无法对单个对象集或属性集进行概念认知学习,而文献[16-17,29]无法对不匹配的对象集和属性集进行概念认知学习.因此,本文提出的概念渐进式认知方法,其适用范围更广.

5 数值实验

本节通过数值实验评估概念渐进式认知算法的有效性.主要包括两个方面:一是本文概念渐进式认知算法自身的评估(即,各种参数对概念认知效果的影响);二是本文概念渐进式认知算法与现有各种概念认知方法的对比.

5.1 实验环境

实验采用的配置如下:CPU为Intel(R) Core(TM) i5-2400 CPU@3.10 GHz,4.00 GB内存,500 GB硬盘;JDK为jdk1.8.0_20,Eclipse使用32位的eclipse-4.2.数据集均来源于机器学习数据库中的真实数据^①,具体如表2所示.

表 2 数据集及其参数

数据集	对象数	属性类型(数量)
Balance Scale	625	离散(4)
Car Evaluation	1728	离散(6)
Nursery	12960	离散(8)
Bank Marketing Data Set	45211	离散(17)
Kddcup. data_10_percent	494021	离散(10),连续(32)
SUSY	5000000	离散(18)

将表2中的原始数据集通过尺度化变换(Scaling)^[31]转化为标准数据集(即形式背景),并分别记为Data sets 1,2,3,4,5和6.如表3所示.

表 3 预处理后的数据集

数据集	对象数	属性数
Data set 1	625	20
Data set 2	1728	21
Data set 3	12960	27
Data set 4	45211	74
Data set 5	494021	163
Data set 6	5000000	38

5.2 实验结果

下面,就线索为对象集、属性集、对象集和属性集的概念渐进式认知问题,分别进行仿真实验分析.

5.2.1 线索为对象集的概念渐进式认知

依据3.1小节,以对象集为线索时,从对象集与步长两个角度评估概念渐进式认知效果.图3(a)至图3(f)分别显示的是给定对象数为3、对象集(即数

据集)为10%及20%时,学习步长变化(步长变化为属性集的10%、20%、30%、40%)的概念认知效果.图3(a)至图3(f)共同表示出认知效果受对象集与认知步长的双重影响.具体地,(1)认知步长一定时,对象集越大,则认知时间越长;(2)认知对象集一定时,认知步长越长,则需要的认知时间越少.这两者与现实中人类认知事物十分吻合,即认知能力一定,对象集越复杂,则认知时间越长;而对对象集一定时,认知能力越强,则所需时间越少.

图4(a)至图4(b)展示的是步长一定时(在此,步长为属性集的10%),对象集变化(对象集的1%、5%、10%、15%)的概念认知效果.图4(a)至图4(b)显示出随着对象增多,则认知时间呈增加趋势的现象.

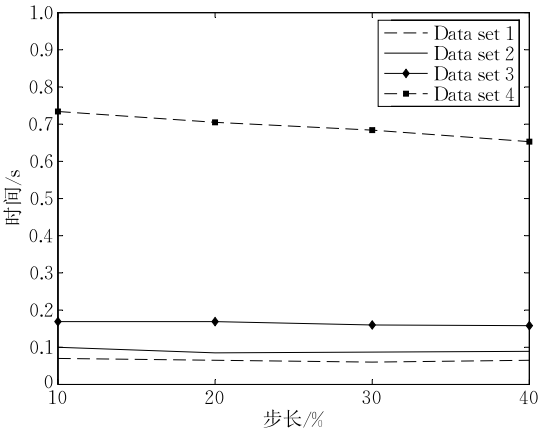
5.2.2 线索为属性集的概念渐进式认知

由3.2小节,取截断图5的参数为:认知步长为2,属性个数分别为3、6、9、12.由图5可知,在认知步长一定时,概念渐进式认知分别受对象集与属性集两方面的影响.具体而言,(1)当属性集一定时,需要认知的对象越多,则所需时间越长,这与人类对越为复杂的事物认知所需时间越长是类似的;(2)当对象集一定时,认知时间随着属性个数的增大而减少,这与现实中人们对属性越为详细的事物识别的速度越快是一致的.另外,图5还表明当对象集较小时(如Data sets 1,2,3),它们之间进行概念渐进式认知所需时间相差不大;而当对象集较大时(如Data sets 4,5,6),它们之间进行概念认知所需时间的差距较大.这说明对象集对概念渐进式认知的影响比属性集大,与3.2小节中将线索为属性集的认知问题归结为对象集的情形是一致的.

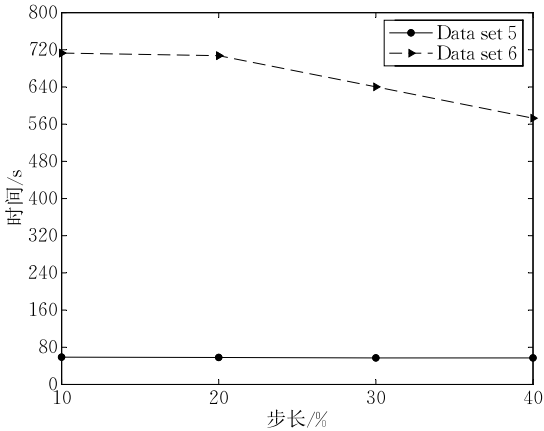
5.2.3 线索为对象集和属性集的概念渐进式认知

根据3.3小节给出的结论,图6是取步长为2,对象集为10%,属性个数的变化为3、6、9、12时悲观取交(图6中,以Data set 1_I,Data set 2_I,Data set 3_I,Data set 4_I,Data set 5_I,Data set 6_I表示)与乐观取并(以Data set 1_U,Data set 2_U,Data set 3_U,Data set 4_U,Data set 5_U,Data set 6_U表示)的概念认知效果.由图6可知,(1)随着数据集不断增大,乐观取并的认知时间比悲观取交要大,且趋于稳定状态;(2)属性数对对象集与属性集同为线索的概念渐进式认知影响较大,即属性越少,则认知时间越长,这类似于人们对属性信息较少

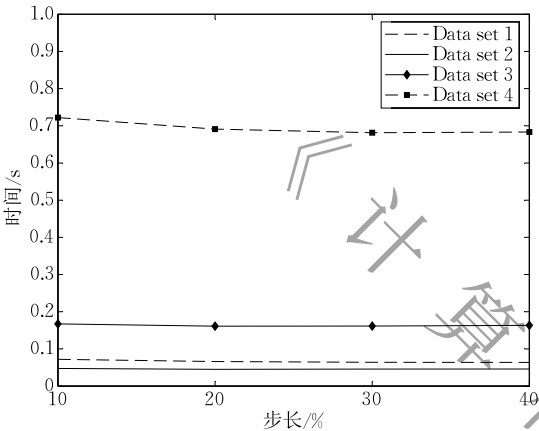
① UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/> 2018.03.15



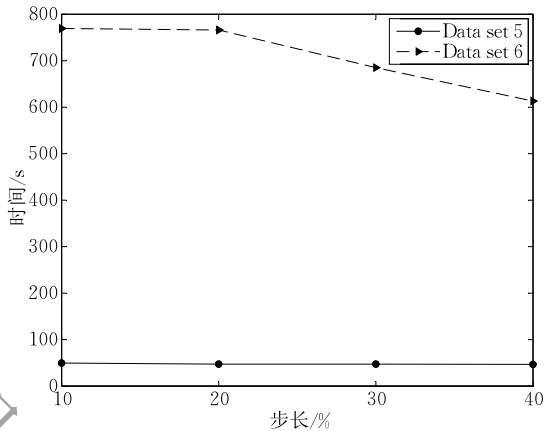
(a) 对象数为3的认知效果



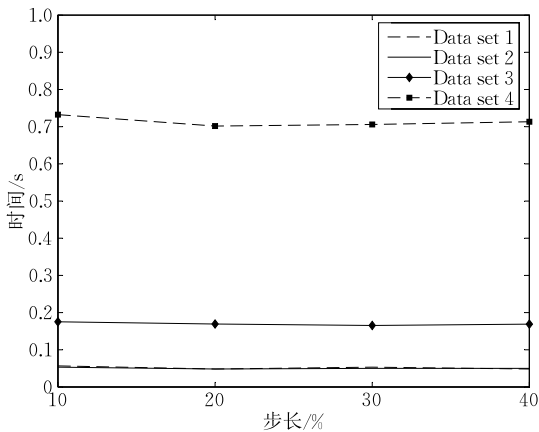
(b) 对象数为3的认知效果(续)



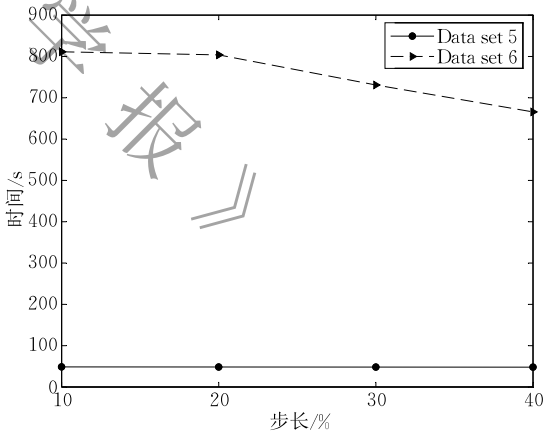
(c) 对象为10%的认知效果



(d) 对象为10%的认知效果(续)



(e) 对象为20%的认知效果



(f) 对象为20%的认知效果(续)

图 3 步长变化时对象集的认知效果

的事物难以快速认知清楚的情况。

表 4 给出的是步长为 2, 对象集大小为 3 的概念匹配程度与认知精度, 其中 E-4 表示“ $\times 10^{-4}$ ”, E-6 表示“ $\times 10^{-6}$ ”。由于每种对象集的具体属性不同, 难以发现对象集之间的精度变化规律。但对于同一对象集, 认知精度随着属性的增加而增加。这与现实中, 信息量增加时概念认知的效果越好是一致的。同时, 表 4 还表明同一对象集匹配程度越高则对应

的认知精度越高; 这同人类对出现的信息越一致的概念认知效果越好的情况是类似的。

表 4 概念认知精度

数据集	属性集基数为 3		属性集基数为 5	
	匹配度	精度	匹配度	精度
Data set 1	0.1429	0.0098	0.3333	0.0102
Data set 2	0.0131	0.0982	0.1176	0.5926
Data set 3	0.0139	0.3333	0.0417	1.0000
Data set 4	3.2737E-4	0.1874	0.0017	0.9629
Data set 5	8.7357E-6	0.0056	8.9784E-6	0.0057
Data set 6	1.1997E-6	1.0000	1.1997E-6	1.0000

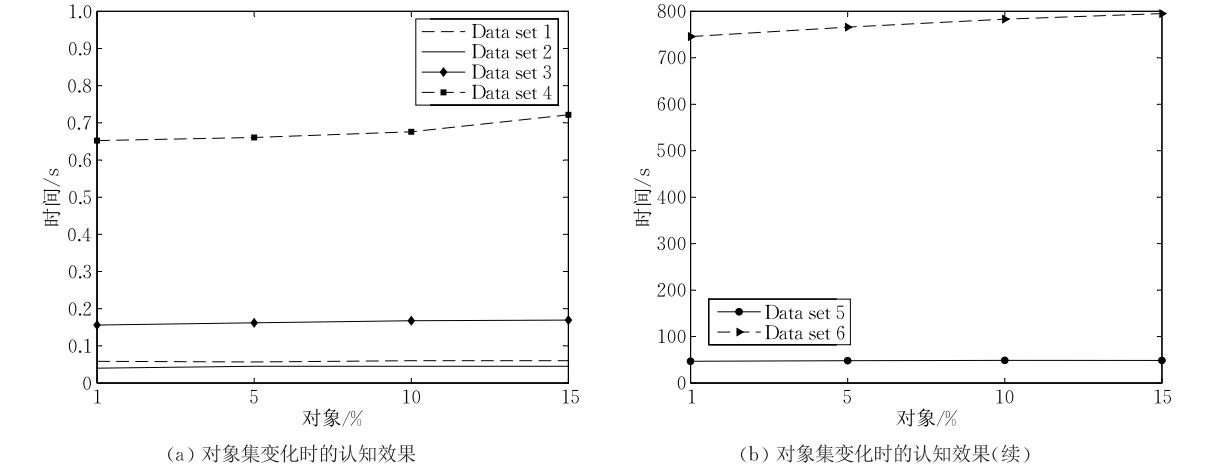


图 4 步长不变时对象集的认知效果

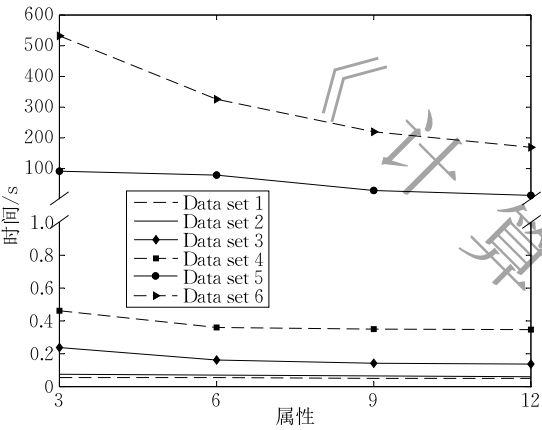


图 5 步长不变时属性集的认知效果

5.3 对比分析

该小节中,将本文的概念渐进式认知方法与文献[17,28]中的概念认知方法进行比较,以进一步说明其有效性.由于本文与文献[17,28]均能够对线索为对象集和属性集的情形进行概念认知,因此下文将从线索为对象集(同时结合属性集信息)的角度比较这三种方法.为了方便,不妨设实验环境中的对象集大小均为 3,文献[28]需要结合属性集大小为 3 的信息进行概念学习,而本文与文献[17]的属性集大小信息不被指定,但以步长为属性集的 40%进行概念渐进式学习.

截断图7给出了本文与文献[17,28]的概念认

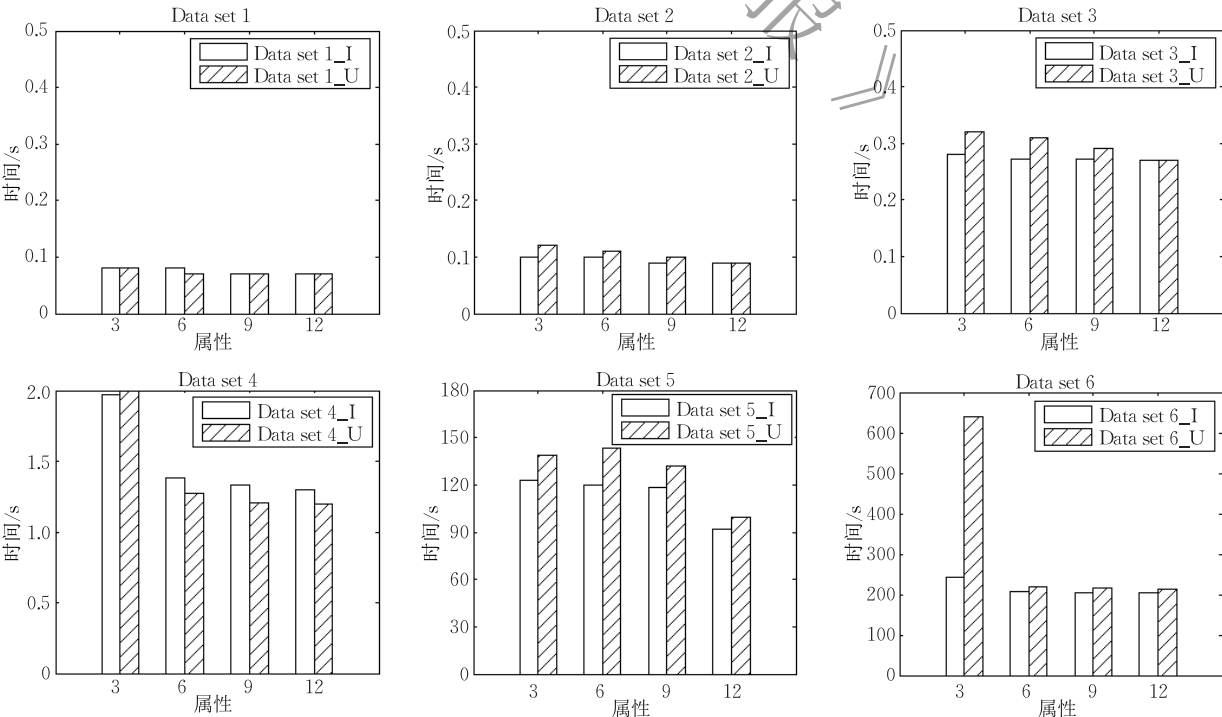


图 6 对象集与属性集同为线索的认知效果

知方法随数据集大小变化呈现出的耗时趋势。由图 7 可以看出, (1) 三种方法的运行时间均随数据集的增大而增加; (2) 本文方法耗时的增大趋势较现有方法有所放缓。另外, 表 5 给出了三种方法的概念认知精度的对比情况, 其中“—”表示文献[28]的学习结果只有单侧概念, 无精度信息。从表 5 可以看出, 文献[28]需要从给定的对象集(同时结合属性集信息)经过充分/必要认知, 获得一个充分必要概念。因此, 尽管最初给定的对象集与本文和文献[17]均相同, 但是进一步结合属性集的信息, 从对象与属性两方面的不断相互转化, 最终所获得的概念与本文和文献[17]均不同。实验结果还表明, 给定对象集, 本文比文献[17]的概念认知精度要高。具体地,

当文献[17]要求的对象属性信息匹配时, 本文与文献[17]的精度相同; 否则, 本文方法的精度高。

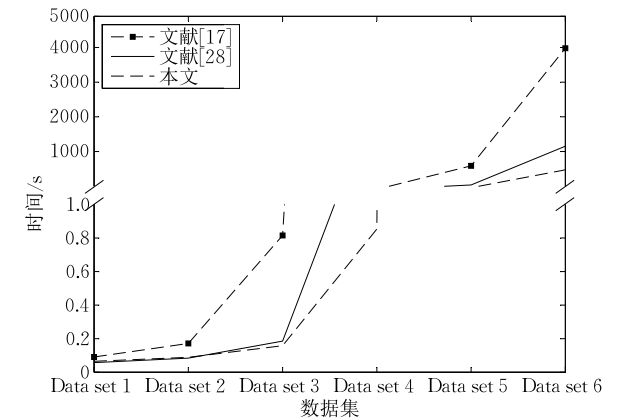


图 7 算法运行时间对比分析

表 5 概念认知精度对比分析

数据集	步长为 40% 的概念认知精度		属性集基数为 3 的概念认知精度		概念认知学习结果
	本文	文献[17]	文献[28]		
Data set 1	1	1	—		本文与文献[17]结果相同， 与文献[28]结果不同
Data set 2	1	1	—		
Data set 3	1	1	—		
Data set 4	1	1	—		
Data set 5	0.0057	对象属性无法匹配	—		本文与文献[17,28]的结果均不同
Data set 6	1	对象属性无法匹配	—		

6 结 论

本文在不完全认知条件下, 通过引入渐进式累积认知函数, 对概念渐进式认知问题做了详细研究。具体地, 提出概念渐进式认知的基本原理, 给出线索为对象集、属性集、对象集和属性集的概念渐进式认知方法, 并与现有概念认知方法进行对比分析, 且仿真实验表明本文方法有效。

相对于已有概念认知研究工作, 本文进一步得到以下结论:

- (1) 概念的渐进式认知, 体现了人们认识事物是一个从不完全到完全的逐步完善过程, 它在某种程度上承认概念认知与时间、空间、环境等因素有关;
- (2) 概念的渐进式认知, 充分考虑了个体偏好、先验知识、背景知识等复杂因素, 符合现实中概念认知的多样性;
- (3) 概念的渐进式认知, 能够实现阶段性认知, 并根据阶段性认知来及时指导下一步的行动, 并逐渐实现完全认知;
- (4) 概念的渐进式认知, 可以认为是模拟人脑进行概念认知的方法, 对概念学习有一定的借鉴作

用。实际上, 一旦人们找到了不完全认知问题的渐进式认知模式, 就可以借助该认知模式, 进一步指导或训练大脑进行有针对性的概念学习;

(5) 概念的渐进式认知, 采用了从粗粒度到细粒度的认知方法, 属于多粒度认知范畴, 故有利于将概念认知逐步优化到最优粒度上。

需要指出的是, 尽管本文在不完全认知条件下对概念渐进式认知问题做了初步探讨, 然而现实中人们认识概念的模式千差万别, 且受到各种不确定因素的影响, 并不能通过一般的抽象化或理想化的认知模型进行完全刻画。比如, 对于多认知主体进行渐进式概念认知问题, 认知主体之间会相互影响, 因为每个认知主体都希望借鉴其它认知主体所发现的概念信息, 以弥补自己认知概念时存在的不足。这是不同空间的概念演化趋同问题。此外, 大脑的认知过程, 除了受渐进式认知影响, 还受大脑自身内部机理环境的制约, 情况非常复杂。因此, 本文提出的几种概念渐进式认知方法, 离实际应用仍有不小差距, 有待进一步完善。

本文的主要研究目的为阐述概念渐进式认知问题的研究价值, 表明它是一个非常复杂的科学问题, 且存在诸多挑战。比如, 大数据环境下的概念渐进式

认知,信息融合或并行计算意义下的概念渐进式认知^[29,32-33],必然属性分析角度下的概念渐进式认知^[34],概念渐进式认知的衡量指标,以及如何将粗糙逻辑推理^[35]等方法应用于概念渐进式认知,以提高概念认知的有效性。

致 谢 西安交通大学张文修教授在本文撰写过程中给予了指导,提出了宝贵的意见和建议,在此表示感谢!

参 考 文 献

- [1] Wang Y X. On concept algebra: A denotational mathematical structure for knowledge and software modelling. *International Journal of Cognitive Informatics and Natural Intelligence*, 2008, 2(2): 1-19
- [2] Wei Ling, Qi Jian-Jun, Zhang Wen-Xiu. Attribute reduction theory of concept lattice based on decision formal contexts. *Science China: Information Sciences*, 2008, 38(2): 195-208 (in Chinese)
(魏玲, 祁建军, 张文修. 决策形式背景的概念格属性约简. *中国科学: 信息科学*, 2008, 38(2): 195-208)
- [3] Ma Yuan, Ma Wen-Sheng. Construction of multi-attributes decrement for concept lattice. *Journal of Software*, 2015, 26(12): 3162-3173 (in Chinese)
(马垣, 马文胜. 概念格多属性渐减式构造. *软件学报*, 2015, 26(12): 3162-3173)
- [4] Liu Zong-Tian, Qiang Yu, Zhou Wen, et al. A fuzzy concept lattice model and its incremental construction algorithm. *Chinese Journal of Computers*, 2007, 30(2): 184-188 (in Chinese)
(刘宗田, 强宇, 周文等. 一种模糊概念格模型及其渐进式构造算法. *计算机学报*, 2007, 30(2): 184-188)
- [5] Li Jin-Jin, Zhang Yan-Lan, Wu Wei-Zhi, et al. Attribute reduction for formal context and consistent decision formal context and concept lattice generation. *Chinese Journal of Computers*, 2014, 37(8): 1768-1774 (in Chinese)
(李进金, 张燕兰, 吴伟志等. 形式背景与协调决策形式背景属性约简与概念格生成. *计算机学报*, 2014, 37(8): 1768-1774)
- [6] Lake B M, Salakhutdinov R, Tenenbaum J B. Human-level concept learning through probabilistic program induction. *Science*, 2015, 350(6266): 1332-1338
- [7] Kumar C A, Ishwarya M S, Loo C K. Formal concept analysis approach to cognitive functionalities of bidirectional associative memory. *Biologically Inspired Cognitive Architectures*, 2015, 12: 20-33
- [8] Wu W Z, Leung Y, Mi J S. Granular computing and knowledge reduction in formal contexts. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21(10): 1461-1474
- [9] Zhang Qing-Hua, Xing Yu-Ke, Wang Guo-Yin. Transformation between a concept knowledge granule and a concept information granule. *Journal of Shandong University (Natural Science)*, 2010, 45(9): 1-6 (in Chinese)
(张清华, 幸禹可, 王国胤. 概念知识粒与概念信息粒的相互转化. *山东大学学报(理学版)*, 2010, 45(9): 1-6)
- [10] Wang G Y, Xu C L, Li D Y. Generic normal cloud model. *Information Sciences*, 2014, 280: 1-15
- [11] Yao Y Y. Interpreting concept learning in cognitive informatics and granular computing. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2009, 39(4): 855-866
- [12] Xu W H, Li W T. Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets. *IEEE Transactions on Cybernetics*, 2016, 46(2): 366-379
- [13] Qiu Guo-Fang, Ma Jian-Min, Yang Hong-Zhi, et al. A mathematical model for concept granular computing systems. *Science China: Information Sciences*, 2009, 39(12): 1239-1247 (in Chinese)
(仇国芳, 马建敏, 杨宏志等. 概念粒计算系统的数学模型. *中国科学: 信息科学*, 2009, 39(12): 1239-1247)
- [14] Liang Ji-Ye, Qian Yu-Hua, Li De-Yu, et al. Theory and method of granular computing for big data mining. *Science China: Information Sciences*, 2015, 45(11): 1355-1369 (in Chinese)
(梁吉业, 钱宇华, 李德玉等. 大数据挖掘的粒计算理论与方法. *中国科学: 信息科学*, 2015, 45(11): 1355-1369)
- [15] Xu Ji, Wang Guo-Yin, Yu Hong. Review of big data processing based on granular computing. *Chinese Journal of Computers*, 2015, 38(8): 1497-1517 (in Chinese)
(徐计, 王国胤, 于洪. 基于粒计算的大数据处理. *计算机学报*, 2015, 38(8): 1497-1517)
- [16] Li J H, Huang C C, Qi J J, et al. Three-way cognitive concept learning via multi-granularity. *Information Sciences*, 2017, 378: 244-263
- [17] Li J H, Mei C L, Xu W H, et al. Concept learning via granular computing: A cognitive viewpoint. *Information Sciences*, 2015, 298: 447-467
- [18] Zhong N, Yau S S, Ma J H, et al. Brain informatics-based big data and the wisdom web of things. *IEEE Intelligent Systems*, 2015, 30(5): 2-7
- [19] Deng W H, Wang G Y, Xu J. Piecewise two-dimensional normal cloud representation for time-series data mining. *Information Sciences*, 2016, 374: 32-50
- [20] Yao T, Kong X W, Fu H Y, et al. Supervised coarse-to-fine semantic hashing for cross-media retrieval. *Digital Signal Processing*, 2017, 63: 135-144
- [21] He K, Mo D Z, Ye T, et al. A coarse-to-fine quasi-physical optimization method for solving the circle packing problem with equilibrium constraints. *Computers & Industrial Engineering*, 2013, 66(4): 1049-1060

[22] Wang G Y, Yang J, Xu J. Granular computing: From granularity optimization to multi-granularity joint problem solving. *Granular Computing*, 2017, 2(3): 105-120

[23] Wang G Y, Xu J. Granular computing with multiple granular layers for brain big data processing. *Brain Informatics*, 2014, 1(1-4): 1-10

[24] Wang Guo-Yin, Zhang Qing-Hua, Ma Xi-Ao, et al. Granular computing models for knowledge uncertainty. *Journal of Software*, 2011, 24(4): 676-694(in Chinese)
(王国胤, 张清华, 马希骛等. 知识不确定性问题的粒计算模型. *软件学报*, 2011, 24(4): 676-694)

[25] Miao Duo-Qian, Zhang Qing-Hua, Qian Yu-Hua, et al. From human intelligence to machine implementation model: theories and applications based on granular computing. *CAAI Transactions on Intelligent Systems*, 2016, 11(6): 743-757 (in Chinese)
(苗夺谦, 张清华, 钱宇华等. 从人类智能到机器实现模型——粒计算理论与方法. *智能系统学报*, 2016, 11(6): 743-757)

[26] Miao Duo-Qian, Xu Fei-Fei, Yao Yi-Yu, et al. Set-theoretic formulation of granular computing. *Chinese Journal of Computers*, 2012, 35(2): 351-363(in Chinese)
(苗夺谦, 徐菲菲, 姚一豫等. 粒计算的集合论描述. *计算机学报*, 2012, 35(2): 351-363)

[27] Zhang Yan-Ping, Zhang Ling, Wu Tao. The representation of different granular worlds: A quotient space. *Chinese Journal of Computers*, 2004, 27(3): 238-333(in Chinese)
(张燕平, 张铃, 吴涛. 不同粒度世界的描述法——商空间法. *计算机学报*, 2004, 27(3): 238-333)

[28] Xu W H, Pang J Z, Luo S Q. A novel cognitive system model and approach to transformation of information granules. *International Journal of Approximate Reasoning*, 2014, 55(3): 853-866

[29] Huang C C, Li J H, Mei C L, et al. Three-way concept learning based on cognitive operators: An information fusion viewpoint. *International Journal of Approximate Reasoning*, 2017, 83: 218-242

[30] King D B, Wertheimer M. *Max Wertheimer and Gestalt Theory*. Piscataway: Transaction Publishers, 2004

[31] Wille R. Restructuring lattice theory: An approach based on hierarchies of concepts//Rival I, ed. *Ordered Sets*. Dordrecht, Holland: D. Reidel Publishing Company, 1982: 445-470

[32] Zhang Tao, Bai Dong-Hui, Li Hui. Parallel concepts computing based on bottom-up decomposition of attribute topology. *Journal of Software*, 2017, 28(12): 3129-3145(in Chinese)
(张涛, 白冬辉, 李慧. 属性拓扑的并行概念计算算法. *软件学报*, 2017, 28(12): 3129-3145)

[33] Mi Yun-Long, Li Jin-Hai, Liu Wen-Qi, et al. Research on granular concept cognitive learning system under MapReduce framework. *Acta Electronica Sinica*, 2018, 46(2): 289-297 (in Chinese)
(米允龙, 李金海, 刘文奇等. MapReduce 框架下的粒概念认知学习系统研究. *电子学报*, 2018, 46(2): 289-297)

[34] Zhi Hui-Lai, Li Jin-Hai. Granule description based on necessary attribute analysis. *Chinese Journal of Computers*, 2018, 41(12): 2702-2719(in Chinese)
(智慧来, 李金海. 基于必然属性分析的粒描述. *计算机学报*, 2018, 41(12): 2702-2719)

[35] She Yan-Hong, He Xiao-Li. Borel probabilistic rough truth degree of formulae in rough logic. *Journal of Software*, 2014, 25(5): 970-983(in Chinese)
(折延宏, 贺晓丽. 粗糙逻辑中公式的 Borel 型概率粗糙真度. *软件学报*, 2014, 25(5): 970-983)



LI Jin-Hai, Ph. D. , professor, Ph. D. supervisor. His current research interests include concept lattice, rough set, granular computing, etc.

MI Yun-Long, Ph. D. candidate. His current research interests include data mining, neural network, formal concept analysis, etc.

LIU Wen-Qi, M. S. , professor, Ph. D. supervisor. His current research interests include data mining, data quality control, decision-making analysis, etc.

Background

In formal concept analysis, concept lattice construction and understanding are two basic problems. After 37 years of development, these two problems have obtained many results which are only applicable for small and middle data sets. Nowadays, however, we have encountered big data in many

research fields. Considering that concept lattice construction is a NP-hard problem, it is impossible to build a concept lattice under the big data environment, let alone well understanding the meaning of the nodes of the concept lattice. Motivated by these problems, in recent years, some experts

(e. g. Zhang Wen-Xiu, Yao Yi-Yu, Wang Ying-Xu, Xu Wei-Hua) concerned concept-cognitive learning (CCL) in which the basic problems include cognitive mechanism of concept-forming, construction and optimization of concept-cognitive system, and simulation of cognitive processes. The main purpose of concept-cognitive learning is to learn the required concepts from a given clue. However, to the best of our knowledge, the existing studies on concept-cognitive learning are based on complete cognitive environment, and they will become ineffective in incomplete cognitive environment. In the real world, incomplete cognitive environment appears more frequently than complete cognitive environment. So, it is crucial to develop concept-cognitive learning models under incomplete cognitive environment.

In this paper, we put forward theories and methods for incremental cognition of concepts from incomplete cognitive environment. To be specific, incremental concept cognition methods are respectively proposed for three types of clues (i. e., an object set, an attribute set, and a pair of object and attribute sets). Especially, we design two strategies

(i. e. optimistic and pessimistic information fusion) to form approximate cognitive concepts, which can at least guarantee that it obtains an upper concept from a given clue in the form of any pair of object and attribute sets. Moreover, numerical experiments are conducted to evaluate the effectiveness of the proposed incremental cognition algorithms. Our methods can be applied to simulating human brain’s cognitive learning and provide a reference to the further study of concept-cognitive learning.

The paper is partial work of the National Natural Science Foundation of China under Grant Nos. 61562050 and 61573173. These projects try to propose some effective methods for concept formulation, learning and cognition based on parallel computing, granular computing and rough set. Up to now, we have discussed concept formulation, learning and cognition from the perspectives of three-way decisions, multi-granularity, information fusion, MapReduce, cognitive computing, etc. And the results of this paper will be a vital step in studying concept-cognitive learning under incomplete cognition environment.