

# 心理常识概念的表示、获取和分析<sup>\*</sup>

田雯 曹存根 王海涛

(中国科学院计算技术研究所 北京 100080)

**摘要** 常识知识是一类重要的人类知识,对自然语言分析、机器智能研究和自动推理研究等都有重要的意义。本文围绕心理常识,主要讨论与心理相关的常识概念的表示、获取和分析的方法。针对现有的概念模型中的概念主要由手工获取,缺乏自动方法,使得概念获取的非冗余性、一致性不能得到保证的问题,本文提出了获取心理常识概念的基本策略:依据心理学中的心理范畴手工获取心理常识的基础概念及概念间关系;根据属性的心理特征手工得到心理属性和属性间关系;以心理常识基础概念和心理常识属性为语义成分,通过“子类生成规则”自动完成获取和组织心理常识概念的任务。并且通过“子类检查规则”检查和分析新加入概念库的常识概念的冗余性和一致性。

**关键词** 心理常识概念,概念获取,概念表示,概念分析

## Representation, Acquisition and Analysis of Psychological Commonsense Concepts

TIAN Wen CAO Cun-Gen WANG Hai-Tao

(The Key Laboratory of Intelligent Information Processing, Institute of Computing Technology,  
Chinese Academy of Sciences, Beijing 100080)

**Abstract** Psychological commonsense is needed in knowledge-intensive applications, such as natural language processing, machine intelligence and real-world reasoning. In this paper, centred on psychological commonsense knowledge, we introduce a framework of acquiring and analyzing psychological commonsense concepts, since such concepts are actually the foundation on which we finally capture psychological commonsense knowledge. In our framework, we have developed a method for acquiring elementary concepts and their semantic relationships based on psychological categories. We introduce five semantical classes of psychological attributes for psychological commonsense concepts, and discuss how to acquire psychological attributes and their relationships based on these semantical classes. To automatically generate compound psychological commonsense concepts, we introduce a rule-based method that generates compound concepts from elementary ones and their pre-defined psychological attributes. We also present an algorithm for checking the inconsistency and redundancy of the concepts.

**Keywords** Psychological commonsense concept, Concept acquisition, Concept representation, Concept analysis

## 1 介绍

随着常识研究的重要性逐渐被人们认识,诸多学者从不同角度、不同学科(社会科学和自然科学)以及这些学科的交叉领域研究常识<sup>[1,2]</sup>。常识研究和心理学研究的交叉成为了目前一个活跃的领域,出现了很多成果和观点,也出现了很多争论<sup>[3~5]</sup>。

心理常识是一种与人的心理相关的常识知识。本文围绕心理常识,主要讨论与人类心理相关的常识概念的表示、获取和分析的方法。心理常识概念是指与人的心理相关的常识概念,其涉及面很广。为了有效地捕捉到心理常识概念,我们根据认知心理学的理论,建立人的各种心理范畴,然后以心理范畴为基础,获取心理常识概念。例如,“感觉”是一个心理范畴,它包括子范畴“视觉”、“触觉”、“味觉”和“听觉”等。在“感觉”范畴或它的子范畴中,我们可以较清晰地定义心理常识概念和它们的内涵。

以 Lenat 为首的常识研究者 1984 年启动了 Cyc 项目<sup>[1,2,6]</sup>,目的是建立一个庞大的常识知识库。我们知道,以产生式表示的规则实际上是逻辑学的蕴含关系,它是推理的基本形式。逻辑学认为知识是一系列命题,命题间存在推理关

系。Cyc 基于该思路构建知识库。到 1995 年,Cyc 建立了拥有 50 万数量级的断言知识库,断言按其知识领域和逻辑上下文关系分为若干组,每组为一个微理论(microtheory)。微理论中含有若干特定领域的相关断言,如时间、空间和因果关系等。每个领域被从不同角度、假设和粒度去观察,对应若干微理论。1991 年前的 Cyc 基于框架表示,采取一阶谓词演算和框架树作为知识表示方法,之后 Cyc 采用逻辑断言形式,其表示语言称 CycL,实际上是一阶谓词演算语言加上等式演算非单调推理性质,以及某些二阶谓词演算的性质。1994 年时 Cyc 积累了约 100 万条常识。其后 Cycorp 公司成立并投入商业运作,到 2000 年已有常识 160 万条。Cyc 虽取得了可观的成果,但也存在不足。Cyc 试图建立一个万能的自然语言处理系统来弥补领域专家系统的不足,例如帮助医学专家系统辨认患者的年龄的填写错误,这显然涉及到浩瀚无边的常识性知识。自然语言本身丰富多样,如果仅采用一阶谓词加自然语言来描述常识,即使数以千万计的规则也很难包容全部常识。常识的基础是语言性知识,语言知识中包括与语言形式无关的概念性知识和纯语言知识,Cyc 没有将两者区别开,忽视纯语言知识在概念获取中的价值,以及混淆纯语言知识和概念性知识带来的负面影响,使得常识断言的数量膨胀,描述代价极

<sup>\*</sup> 本文得到自然科学基金的支持( #60273019 和 #60037017)、科技部基础研究专项的支持( #2001CCA03000)。田雯 博士研究生 主要研究方向:常识知识获取和分析。曹存根 博士生导师,主要研究方向:人工智能。王海涛 博士研究生,主要研究方向:自动知识获取。

大。故虽其目标定位准确,但不能一蹴而就,需要更复杂的知识表示机制,而不仅仅是常识断言的累加。

盘古计划<sup>[7]</sup>是陆汝钤院士带领进行的“常识性知识的实用研究”项目。该系统的常识知识库分为三个基本层次:应用层,处理各类提问的内容和回答方式;问题类型层,归纳各种提问的内容和方式;问题表示层,使用语义网络表示知识。基于系统功能,确定了常识知识库的组织原则,包括常识知识的模块化、基于 Agent 的专家系统和 Agent 间的通信,对模糊推理、非单调推理的表述等。盘古系统是对常识的实用性的研究,创造性提出了新的封闭世界假设,并可进行模糊推理。该系统偏重于如何利用常识来进行解决实际问题的研究。

我们需要指出,常识知识的获取和表示与语言知识的获取和表示也有密切的关系。WordNet、MindNet 和 FrameNet 等项目对常识概念模型基于语义的研究做出了有益的尝试。MindNet 系统研究如何从词典中自动获取概念知识<sup>[8-9]</sup>。它通过自然语言广域分析器,采用全倒置技术,在概念的解释或例句中寻找概念语义。WordNet 是一个应用广泛的在线英语词库<sup>[10]</sup>。各组同义词之间通过同义关系、反义关系、上位/下位关系、部分/整体关系组成庞大的语义网络,其概念的内涵描述是非形式化的。FrameNet 的主要思想是,由人以机器可读的形式对概念语义编码,采用框架语义描述数千英语词语,并用带手工语义标注信息的语料库给以支持,框架表示使 FrameNet 的概念语义明晰,但手工标注工作量大,缺乏自动获取常识概念的手段<sup>[11]</sup>。

中文的 HowNet 和《同义词词林》从语义学和语法角度总结了中文概念的结构规则,并形式化了若干常识概念。1983 年梅家驹编撰《同义词词林》,实现了汉语词语基于语义的索引,采用词群表示概念的同义词/同类词集合,3925 个词群构成语义树<sup>[12]</sup>,其目的是使用户可在同义词集中找到合适词表达概念,因此获取词群时更多考虑语用,而非严格语义。同义词集合给出了成员概念,但没给出概念的内涵描述。HowNet 是以英汉双语词代表的揭示概念特征及概念间关系的网状常识知识库,着力刻画概念的共性和个性,除刻画词间关系外,更试图描述英语和汉语词语所代表概念间,以及所具有属性间的关系<sup>[13-14]</sup>。

《同义词词林》与 WordNet 的概念均以同义词为单位建构概念库。但 WordNet 更强调词间关系,是基于同义词/反义词来描述概念间关系的语义网络,因其丰富的概念和明确的语义,在应用系统中使用较多。FrameNet 采用简洁的框架表示,并有语料支持,也得到了普遍关注。HowNet 更关注概念间深层的本体论关系,其构建常识库的目标远大,需要走的路也很长。MindNet 采用自动手段获取概念,但由于其语料的结构较固定简单,具有一定局限性。

1981 年 Osherson 和 Smith 提出评价概念模型的两个标准<sup>[15]</sup>。第一是合成性,即须在概念模型中包括组合规则来描述复杂概念如何基于简单概念进行组合。例如设 A、B 是两个概念,需提供组合复杂概念 A-and-B、A-or-B 和 A-B 等规则组合生成复杂概念。第二个标准是能给出能判断概念间明确语义关系的条件,例如给出“所有苹果都是水果”的条件。

前述几个常识概念系统没有给出复杂概念如何组织的说明,概念的获取和组织往往基于知识工程师的个人理解,难以保证概念获取的相对完整性和一致性。并且,除 MindNet 外的其他项目大都由手工完成,缺乏自动方法,因而工作量成为常识概念获取的瓶颈。对第二个标准来说,由自然语言表示的常识概念非常丰富,且常识概念是心理常识断言的基础,心理常识断言的一致性要求心理常识断言中出现的概念须是已定

义、具有形式化表示、处于常识概念结构的某个位置上,并且能确定概念间的明确语义关联的,所以,常识概念须具有相对的完备性。如果仅仅根据知识工程师的脑海的搜索,很难将常识概念和概念间关系列举全面。

基于上述讨论,我们在建立心理常识概念库时,将克服上述问题,通过构造与构词法和人们使用习惯相一致的(子类生成规则)来自动生成复合概念,并用语义清晰的框架来描述概念的语义和概念间的关系,最后,采用与(子类生成规则)一致的(子类检查规则)来对新加入的概念进行检查,保证常识概念库的一致性,并解决冗余性问题。什么是常识概念呢?我们通过常识概念和科学概念之间的区别来理解常识概念的特点,表现在以下几个方面:

(1)模糊性:常识概念常缺乏精确定义。定义的模糊性表现在三个方面:一是概念外延的定义不清,实例对常识概念的隶属关系往往不确定,不是严格属于或不属于,而是含混的,根据语境(如人、时间和场合等)的不同而不同。例如,一个具体的“鲸鱼”对常识概念“鱼”的隶属关系会因不同人对“鱼”和“鲸鱼”的不同理解而不同。二是概念的内涵定义不清,常识概念的属性取值往往不是精确的,而是模糊值,且对属性取值的理解会因人而异,例如,不同语境中人对“年轻”的理解不同。三是常识概念的使用中的模糊性,人们很少去追究如何精确使用常识概念,也很少用科学手段验证其特征,而是在“求同存异”的“共识”中使用常识概念。

(2)构词的规律性和偶然性:常识概念是在长期生产生活中积累词汇形成的,所以常识概念的构词既有认知规律起作用,即“构词法”,也有历史的偶然性或习惯起作用。我们在建立心理常识概念库时,利用构词规律来自动生成心理常识复合概念;

(3)无处不在性:科学概念往往来自于有边界的科学领域。边界性表现在:可有效判定概念是否属于该领域的讨论范围。常识概念涉及生活的各个方面。因此,如何限定常识概念集的边界,如何获取和组织这些“散乱”的常识概念是常识概念获取的关键问题之一。获取心理常识概念的第一步是确定将哪些常识概念加入心理常识概念库。

(4)私有性:不同语境对应不同的常识概念集。例如“病毒”对医务工作者是常识,对小学生是专业词汇;“饺子”是中国文化中的常识概念,而非外国文化中的常识概念;

我们主要获取与人的心理相关的常识概念。心理学的研究者认为,人们头脑里存在概念的“实例”,以及对概念的“实例”的描述,这种描述无可置疑地对应于现实世界的真正实例,同样也存在概念间关系与现实世界中实例间关系的对应<sup>[16]</sup>。例如,苹果和盘子间的空间关系,在头脑中可表示为“苹果放在盘子里”。但即便人们有这样的认识,仍不能完全描述苹果和盘子之间的整个空间关系。因此,人的心理中对事物的描述不等同于实际中的描述。那么人的心理中的概念以一种什么样的方式存在并组织呢?

Palmer 指出<sup>[17]</sup>,人的心理对外界的认识是大脑对现实事物及其关联通过两次提取得到的。该理论表述为:设  $R_1$  是人的心理, $R_2$  是人的头脑对现实世界 D 的认识,则  $R_1$  是人对  $R_2$  的心理反应,而不是对 D 的直接心理反应。根据该理论,我们在建立心理常识概念库时,试图提供这样的概念获取和组织方法,它不关注现实世界本身是如何组织的,而更关心人们心理中概念的组织方式,并更多考虑概念的心理特点。我们的方法是,以人的心理范畴为依据获取心理常识概念,不同的心理范畴对应于不同的心理常识概念及结构,而后整体构成心理常识概念库。本文获取、表示和分析心理常识概念的步骤包

括:

- (1)以心理学本体中的心理范畴为依据,获取心理常识基础概念,及其语义框架;
- (2)获取心理常识属性集,使用属性框架表示其语义;
- (3)自动生成心理常识复合概念,采用框架表示其语义;
- (4)在心理常识概念生成的同时,组织所有概念为按心理范畴划分的心理常识概念库;
- (5)对新加入心理常识库的常识概念,判定它与现有心理常识概念库的一致性和冗余性。

本文第2节讨论心理常识概念的语义成分,包括心理常识属性和心理常识基础概念的获取和表示,为自动生成心理常识复合概念做准备;第3节介绍心理常识复合概念的获取方法;第4节分析新加入的常识概念与现有心理常识概念库的一致性和冗余性;最后总结全文。

## 2 心理常识概念的语义成分

在心理常识概念中,概念可分为基础概念和复合概念。基础概念是指原子概念,在语法和语义上,均不能将其划分为各个独立成分,如苹果、水等。基础概念的获取依靠手工。复合概念可被分为独立的语义成分,各个语义成分以两种形式出现在心理常识复合概念的生成规则中,心理常识基础概念和心理常识属性的取值。我们先从语义的角度,分别介绍修饰物理实体的语义成分和修饰抽象事件的语义成分,然后具体讨论各语义成分。

从语义上看,物理实体所对应的心理常识概念的语义成分包括:

- (1)涉及到的人:如“使用者”、以“性别”、“年龄”等划分的人群或“发明者”等;
- (2)存在和来源:包括在时间意义上的存在和来源,如“朝代”、“出生时间”等;在空间意义上的存在和来源,如“座落地”、“经度”、“出生地”等;在社会意义上的存在和来源,如“所属国”、“民族”、“社会阶层”和“宗教”等;
- (3)人的心理上的特征:如感觉特征、知觉特征、记忆特征等。感觉特征包括视觉特征(如“颜色”和“高度”)、听觉特征(如“音高”)、味觉特征(如“食味”)和触觉特征(如“粗糙度”)等,知觉特征包括概念是整体还是部分、“美感”等;
- (4)人工物的制法:例如“食品”的“烹调手段”、“材料”和“生产工艺”等;
- (5)功能:用途,如“衣服”和“鞋子”的“功能”。
- (6)成分和含量:对某种成分的含量大小。例如食品中的“糖含量”和“纤维含量”等;
- (7)喻象或者象征意义等:如“梅花”的“象征意义”。

抽象事件对应的心理常识复合概念的语义成分通常有:

- (1)知觉属性:包括时间特征(如“发生时间”)、空间特征(如“发生地”);
- (2)涉及到的人和事物:如“创始人”、“领导人”、娱乐事件“玩篮球”中的“球类”;
- (3)事件的动作:如“玩”篮球;
- (4)事件的原因,如历史事件的“导火索”和“历史起源”等;
- (5)事件的结果,如历史事件的“影响力”和“死伤人数”等。

### 2.1 心理常识属性的获取

2.1.1 心理常识属性的分类 心理常识属性是心理常识概念的属性及取值,是构成心理常识复合概念的重要语义成分。为方便心理常识属性的获取和组织,以及今后的推理工

作,根据心理范畴将心理常识属性分为:

(1)依赖感觉的属性(sense-dependent):人们要认识该类属性的值,须具有相应感觉。例如“感觉属性”包括“可视属性”和“可听属性”等,“可视属性”又包括“颜色”属性,盲人不具备视觉,因而不能正确识别事物的“颜色”。

(2)依赖教育的属性(education-dependent):人们要认识该类属性的值,需借助专门知识的学习和研究。例如:疾病的“病因”是“病毒”。

(3)依赖思维的属性(mind-dependent):人们要认识该类属性的值,需要借助思考。例如:某个历史事件的“个人观点”。

(4)依赖想象的属性(imagination-dependent):人们要认识该类属性的值,需借助想象。例如,对某个书中情节或者人物的描述,事物的“象征意义”等。

(5)依赖经验的属性(experience-dependent):人们要认识该类属性的值,需要有过相应的经验,例如对“人”的“身份”的理解需主观上了解该身份的含义。经验的获得源于“感觉”“教育”等,故依赖经验的属性一定是依赖感觉、教育、思维、或想象的属性。

以上从整体上对心理常识属性进行分类,在我们的系统中,心理常识属性之间的关系有两种:父子属性之间的关系,以及父集合属性和其成员属性间的关系。

定义1(父集合属性和成员属性) 设 $c$ 是常识概念, $A$ 是常识属性,其值域为 $D_A$ ,如果 $A$ 是常识属性集合 $\{B_1, B_2, \dots, B_n\}$ 的父集合属性,当且仅当对任何属性值 $v \in D_A$ ,如果 $value(c, A) = v$ ,则至少存在一个 $B_i$ ,使得 $value(c, B_i) = v$ 成立。 $\{B_1, B_2, \dots, B_n\}$ 称为 $A$ 的成员属性集合,其中每一个属性 $B_i$ 称为 $A$ 的成员属性。

定义2(父属性和子属性) 设 $c$ 是常识概念, $A, B$ 是常识属性, $D_B$ 是 $B$ 的值域。 $A$ 是 $B$ 的父属性,当且仅当对任何属性值 $v \in D_B$ ,若 $value(c, B) = v$ ,则 $value(c, A) = v$ 也成立。也称 $B$ 为 $A$ 的子属性。

上述两种关系不相同,父子属性之间的关系是自然形成的,在实际中有广泛的应用,在知识工程中研究也很多。例如,“颜色”和“色调”之间是父子属性。而“可视属性”和“颜色”属性之间构成父集合属性和子属性类之间的关系,“知觉属性”和“美感”之间是父集合属性和成员属性的关系。我们人为地把父集合属性引入我们的系统中是为了:

(1)描述人们要了解某心理常识概念的属性值,在心理上需具备的必要条件,例如“颜色”是“可视属性”的成员属性,则它依赖于视觉,没有视觉能力的人不能感觉到该属性。通过对属性心理条件的表述,实现了对常识概念“私有性”的描述,概念之所以成为某心理语境下的心理常识概念,是因为该语境具有认识该概念的属性的心理条件。

(2)方便心理常识属性的归纳,以心理范畴(如感觉、知觉)来分类心理常识属性有助于人们理解心理常识属性在心理上的特征。

(3)在我们的系统中,心理常识概念也以心理范畴为依据进行获取,所以,以属性在心理上的特点建立心理常识属性的分类体系,将为获取心理常识概念提供基础和方便。

2.1.2 心理常识属性值 除了常识属性之间具有关系外,属性取值之间也存在联系。根据属性取值的模糊性和结构特征,将心理常识属性值划分为:

(1)结构化属性值,包括数值属性值、模糊属性值和分类属性值。A)数值属性值:属性值是数量值。如:“年龄”为“55岁”;B)模糊区间属性值:可根据语义进一步量化为区间数量值的模糊语义字符串,如“年龄”的属性值为“成年”;C)分类

属性值:属性取值间具有偏序关系,属性值之间构成层次结构,例如“人”的“从事领域”属性值域中包括化学、有机化学、物理、高能物理等,属性值构成偏序的层次关系,包括:“有机化学是一种化学”,“高能物理是一种物理”等。

(2)非结构化属性值:也称离散属性值,各属性值互不关联,例如“商场”的“经营商品”包括:“水果”、“衣服”和“音像制品”,且属性值不能数值化和区间化。

我们重点讨论分类属性值,因为它是心理常识复合概念的生成规则的重要语义部分之一,其特殊性在于属性值之间存在偏序关系,使得生成的复合概念间具有某种层次关系。另外值得注意的是,常识属性的取值可能是某常识概念,如“人”的“职业”是“化学家”,该值是常识概念,在概念库中有相应的框架描述,由于概念框架有对其子概念的描述,所以分类属性值的值域对应了以层次关系组织的心理常识概念集。为讨论

方便,设每个常识属性值的根结点为空结点(记为 $\lambda$ ),常识属性值间的层次关系记为“ $\angle_v$ ”,形式化定义如下:

定义3(属性值的蕴含关系) 设心理常识属性A的值域为 $D_A$ 。属性值 $v, v' \in D_A$ 。对于任意心理常识概念c,如果c在A上的取值是v,蕴含了c在A上的取值是v',称v蕴含了v',记做 $v \angle_v v'$ 。

定义4(分类属性) 设心理常识属性A的值域为 $D_A$ ,最顶端属性值 $\lambda$ 为空属性值,A是分类属性,当且仅当对于偏序关系“ $\angle_v$ ”,存在 $v_1, v_2 \in D_A, v_1, v_2 \neq \lambda$ ,有 $v_1 \angle_v v_2$ 或 $v_2 \angle_v v_1$ 成立。

定义5(直接父属性值) 属性值 $v_1$ 是属性值 $v_2$ 的直接父属性值,则 $v_2 \angle_v v_1$ 成立,且对于任意 $v_2 \angle_v v_3$ ,有 $v_1 \angle_v v_3$ 成立。记为 $v_1 = sup(v_2)$ ,或 $v_2 = sub(v_1)$ 。

例1 图1是分类属性的简单示例。

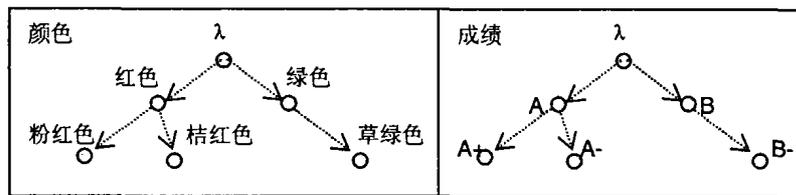


图1 分类属性示例

如图1,“颜色”是分类属性,“粉红色”和“红色”间存在蕴含关系。“成绩”是分类属性,其值域是{A, A+, A-, B, B+, B-, C, C+, C-, D, D+, D-},“A+”和“A”之间存在蕴含关系。

2.1.3 常识属性的选择和描述 描述常识概念的常识属性很多,可以从不同的角度来列举和分类。在选择心理常识属性来组合生成心理常识复合概念时,我们筛选与人的心理有关的心理常识属性来组合生成心理常识复合概念,选择心理常识属性的基本原则为:

(1)选取与人的心理相关的常识属性,相关性指:A)能明显地引起人的某种心理现象,如“食物”的“食味”和“香味”;

B)能反映不同人的心理的差异,例如“人”的“年龄”不同,心理上有很大差异。

(2)心理常识属性之间彼此尽量独立;

(3)心理常识属性的值域空间为以蕴含关系“ $\angle_v$ ”组织的层次结构;

(4)任一属性值对应一个心理常识基础概念;

使用框架表示心理常识属性的语义,框架按语义分为心理常识属性集合(如图2(C)),心理常识属性类(如图2(B))和心理常识属性实例(如图2(D)),其具体描述语言在图2中给出。

<pre>defcategory 槽 {   是子类: 术语   属性: 值域   属性: 类属   属性: 单位   属性: 缺省值   属性: 类型   属性: 同义词 } (A)</pre>	<pre>defcategory &lt;属性类名&gt;: 槽 {   [有父类: &lt;属性类名向量&gt;]   [有父集合: &lt;属性集合名向量&gt;]   [有子类: &lt;属性类名向量&gt;]   [有子槽: &lt;属性名向量&gt;] } (B)</pre>
<pre>defcategory &lt;属性集合名&gt;: 槽 {   [有父集合: &lt;属性集合名向量&gt;]   [有子集合: &lt;属性集合名向量&gt;]   [有子类: &lt;属性类名向量&gt;]   [有成员: &lt;属性名向量&gt;] } (C)</pre>	<pre>definstance &lt;属性名&gt;: 槽 {   [有父集合: &lt;属性集合名向量&gt;]   [有父类: &lt;属性类名向量&gt;] } (D)</pre>

图2 心理常识属性的描述语言说明

图2(A)是对心理常识属性的属性类和属性实例的格式说明,可选择“槽”中定义的项目描述所有属性类和属性实例,“是子类”是一个关系,指明“槽”类是“术语”的子类。“值域”用于指明属性子类或属性实例的值域范围;“类属”用于指明“槽”的子类属性或属性实例所修饰的心理常识概念;“单位”

用于指明的属性类或属性实例的取值单位,如“公里”;“类型”用于指明“槽”的属性类或属性实例的取值的数据类型,如“字符串”、“数量值”等;“缺省值”用于指明当“槽”的属性类或属性实例未被赋值时系统的默认取值。

图2(B)属性类说明。其上位属性有属性类或属性集合,

用“有父类”和“有父集合”表示,其下位属性是属性子类或者属性实例,用“有子类”和“有子槽”表示。图 2(C)属性集合说明。其上位是属性集合,用“有父集合类”表示,下位包括属性子类、属性子集或成员属性,用“有子类”、“有子集合”、“有成

员”表示。图 2(D)属性实例的说明,其上位是属性集合或者属性类,用“有父集合”和“有父类”表示,无下位属性。

例 2 图 3 举例说明属性集合,属性类和属性实例的框架和它们之间的关系。

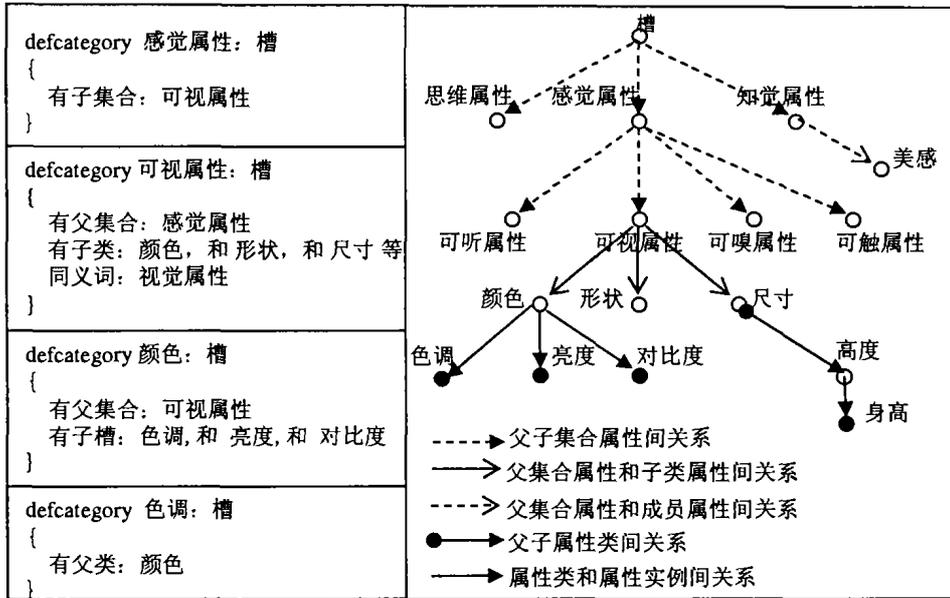


图 3 心理常识属性框架示例

## 2.2 常识基础概念的获取

2.2.1 基本思路 获取心理常识概念的第一步是获取心理常识基本层次概念。常识概念涉及到生活的各个方面,这些概念在不同领域中以不同标准划分,形成不同的概念分类系统(如生物学对生物的分类体系)。心理常识概念库中的概念分类与心理相关。我们依据心理范畴来获取常识基本层次概念。例如图 4 中“视觉”心理范畴将概念以“颜色”等进行分类;根据马斯洛的需求理论,与“需要”范畴相关的概念分类为“食物”、“衣物”等。每个心理范畴对应了对心理对概念的分类体系,各分类按心理范畴组织,构成人的心理中常识概念的组织结构。

例 3 图 4 是按照心理范畴进行概念分类的示例。

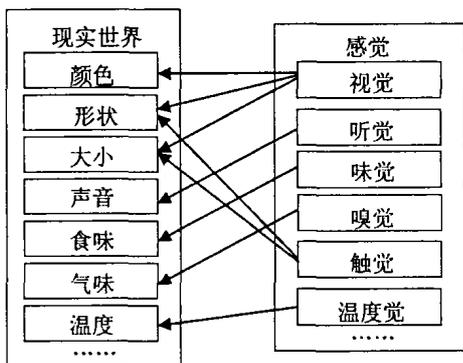


图 4 按心理范畴来划分常识概念示例

2.2.2 常识基础概念的表达 基于常识属性和常识基础层次常识概念,我们通过定义心理常识复合概念的生成规则来自动生成复合常识概念的框架表示(包括概念的名称、属性和关系说明)和常识概念间的层次关系。这里,生成的心理常识复合概念是常识基础概念的子概念,其生成规则在手工获取的基础层次常识概念框架中用“子类生成规则”给出,常识概念的描述语言为图 5。

defcategory(心理常识概念名)

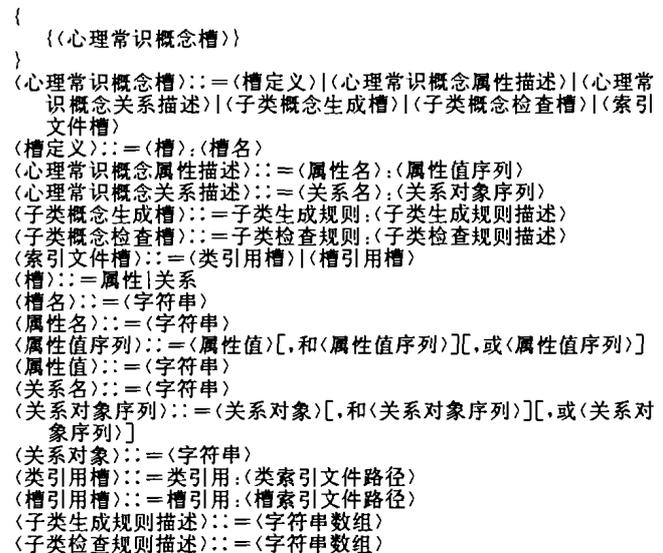


图 5 心理常识概念框架描述语言

(心理常识概念槽)包括六个部分:

(1) (槽定义): 该常识概念的子类或实例的描述可使用的属性和关系,及应遵循的格式;

(2) (常识概念属性描述): 常识概念的属性及值,如“苹果”的“颜色”的取值。该类属性分为两类:可被该概念的子类继承的公用属性;不能被继承的私有属性;

(3) (常识概念关系描述): 当前描述的常识概念与其他常识概念之间的关系,如概念“人”的“有子类”关系的描述,对应概念“儿童”等;

(4) (子类概念生成槽): 生成当前描述的常识概念的子类的规则,所生成的心理常识复合概念是当前描述的心理常识概念的子概念。使用子类生成规则的文法约束“子类生成规则”的语义和书写格式;

(5) (子类概念检查槽): 说明当前描述概念的子概念应遵守的格式,用于检查新加入的常识概念与常识概念库的一致

性或冗余性。使用子类检查规则的文法约束“子类检查规则”的语义和书写格式；

(6)〈子类生成规则索引文件槽〉：用于给出“子类生成规则”和“子类检查规则”中使用的心理常识属性描述和常识基础概念描述所在文件的列表。

例4 图6是心理常识基础概念“人”的语义框架的示例。

```
defcategory 人
{
  英文: person
  是子类: 动物
  类引用: c:\ren\categories-ren.txt, 和 c:\categories-all.txt
  槽引用: c:\ren\slots-ren.txt, 和 c:\slot-all.txt
  子类生成规则: [国家子类]〈人·性别〉\replace(国家子类, 国籍)
  子类检查规则: [国家子类]+[民族人类]+[人·种族]+[人·性别]+人
}
```

图6 心理常识基础概念“人”的语义框架

“英文”是“人”的属性，“person”是其值；“是子类”是概念间父子关系的描述；“子类生成规则”是自动生成复合常识概念的规则。〈子类检查规则〉检查新加入的常识概念。其文法和书写规则在下一节介绍。

### 3 心理常识复合概念

在获取和表示了心理常识基础概念和常识属性的基础上，我们可通过〈子类生成规则〉来自动生成心理常识复合概念，并可以通过〈子类检查规则〉对新加入的概念进行一致性和冗余性的检查。所以，心理常识概念库由心理常识基础概念、心理常识属性、〈子类生成规则〉和〈子类检查规则〉组成。

定义6(心理常识概念库) 一个心理常识概念库  $B = \{Attr, X, F, S\}$  是一个四元组，其中， $Attr = \{A_1, A_2, \dots, A_n\}$  是心理常识属性集合， $X = \{x_1, x_2, \dots, x_n\}$  是心理常识基础概念集合，子类生成规则集合  $F = \{f_1, f_2, \dots, f_p\}$ ，子类检查规则集合  $S = \{s_1, s_2, \dots, s_q\}$ 。

在  $Attr$  和  $X$  的基础上，通过子类生成规则集合  $F$ ，可自动生成并组织心理常识复合概念为心理常识基础概念的子类。子类生成规则集合  $F$  的文法如图7。

```
〈子类生成规则描述〉 ::= 〈子类生成限定〉[〈子类属性替换〉]
〈子类生成限定〉 ::= 〈限定项〉[〈子类生成限定〉]
〈限定项〉 ::= ' [ ' [ 〈基本限定项〉 [ ' ] ] [ ' ] [ 〈组合限定项〉 [ ' ] ] ]
〈子类属性替换〉 ::= replace(〈子类原限定项〉, 〈新子类属性〉) [ ; 〈子类属性替换〉 ]
〈基本限定项〉 ::= 〈关键字限定〉 | 〈槽限定〉 | 〈类限定〉 | 〈侧面限定〉
〈组合限定项〉 ::= ' / ' (〈基本限定项名〉 [ 〈基本限定项名循环〉 ] ) * [ 〈组合限定项〉 ]
〈关键字限定〉 ::= 〈关键字〉
〈槽限定〉 ::= ' / ' (〈槽限定名〉)
〈类限定〉 ::= ' / ' (〈类限定名〉)
〈侧面限定〉 ::= ' / ' (〈侧面限定名〉)
〈基本限定项名〉 ::= 〈关键字〉 | 〈槽限定名〉 | 〈类限定名〉 | 〈侧面限定名〉
〈基本限定项名循环〉 ::= ' / ' (〈基本限定项名〉 [ 〈基本限定项名循环〉 ] ) *
〈槽限定名〉 ::= 〈类名〉. 槽名
〈类限定名〉 ::= 〈类名〉子类
〈侧面限定名〉 ::= 〈类名〉. 〈侧面名〉子类
```

图7 子类生成规则的文法

(1)〈子类生成限定〉是〈限定项〉的有序集；

(2)〈子类属性替换〉：在生成子类框架时，如果用于组合的属性或概念的名称在子类框架中不适宜做为属性出现，则换用该项规定的属性名，如，在“食物”的子类生成槽里有“国家”语义成分，则生成子类“中国食物”的框架中，使用“来源国家”，取值为“中国”。

(3)〈限定项〉是〈基本限定项〉和〈组合限定项〉的有序集；

(4)〈基本限定项〉中的“槽限定”表示复合概念的语义成分是属性值，当属性取值是常识概念时，用于组合的属性值包括该属性值域对应的概念及其所有子概念。“类限定”表示用于组合复合概念是某心理常识基础概念及其子类。“关键字限定”表示组合复合概念的语义成分是字符常量。“侧面限定”表示用于组合复合概念的语义成分是某常识概念的子类集合，且子类集合是当前的常识概念以指定的心理常识属性划分的，例如用“年龄”划分“人”得到各个年龄组的人群“老年人”、“儿童”和“青年”等；

(5)〈组合限定项〉是基本限定项的有序集。

在书写“子类生成规则”和“子类检查规则”时还需考虑各个语义成分的组合顺序。英文语法中，形容词分为限制性形容词和描述性形容词<sup>[18]</sup>。限制性形容词表示事物的本质，位置紧挨所修饰名词，两者的关系如同一体，限制性形容词对生成的复合概念不可缺少，否则会影响名词的意义。描述性形容词又称非限制性形容词，仅起到描绘的作用，其位置在限制性形容词之前，省略时不会影响复合名词的语义。另外一种形容词修饰的顺序标准是“限描大颜内”的构词现象。“限”指限定性形容词，即数量词、物主、不定代词和指示代词等，如“一个”、“我的”；“描”是人们对概念的心理感受的形容词，如“美丽的”等。“大”指“大小”，“新旧”和“形状”等表示事物特征的形容词，如果这些词同时出现，则按上述次序依次接近被修饰概念。“颜”是“颜色”、“味道”等感觉属性。“内”是指概念实例的内在属性，包括“材料”、“化学成分”和“质地”等。根据上述构词规律以及在概念获取中的经验，归纳概念语义成分的组合顺序的规则为：

(1)非本质属性在前，本质属性在后。属性是否本质取决于两个方面：从外延上，本质属性尽可能把概念的实例范围限定得较确定和明确。设概念集为  $U$ ，则本质属性是那些对  $U$  进行划分时可得到尽量不相交的分类的属性，分类后实例对概念的隶属关系越明确，该属性就越本质。例如，任何一个“食物”的实例，如“面包”或“苹果”，人们较易确定它们是否是“面食”的实例，而较难判断它们是否是“中国食物”的实例。从内涵上看，本质属性是指能更明确体现事物特征的属性。如“衣服”的“款式”取值为“夹克”属性值比“所属国家”取值为“中国”，更容易让人们在心理上认识该概念的内涵特征，更容易使人们在脑海中建立概念对应的实例的形象特征；

(2)模糊的属性在前，精确的属性在后。精确性表现在两方面，第一种精确性是指如果人们越容易直观上判断事物是否具有该属性，该属性就越精确。“衣服”中“材料”属性较容易让人们识别，而“产地”就不易直观到。第二种精确性是指属性的取值精确。

(3)限定性在前，描述性在后。与英语语法中形容词的构词规则一致；

(4)不稳定的在前，稳定的在后。属性取值在不同语境中可能不同，较稳定的属性不轻易随语境变化而改变取值，这样的属性可决定概念不易随语境改变的本质特征；

(5)有关社会的属性在前，有关感觉的属性在后：社会属性是指“国籍”、“民族”等和社会相关的属性，表示的是心理常识概念与社会的关系，是人为或社会赋予的，故不如“感觉”等人们直接观察到的属性本质。感觉的属性如“颜色”和“声音”等。

以上心理常识复合概念的语义成分的组合顺序规则是经验的总结。事实上，由于心理常识概念的形成本身有着复杂的

社会历史背景,我们在组合概念时还应该考虑用词习惯,选择自然的、贴近人们习惯的语义成分顺序。为描述的方便,我们假定,空概念是所有概念的父概念,空属性值被所有其它属性值蕴涵。心理常识概念的基本定义如下:

**定义7(心理常识复合概念)** 设  $c$  可分为各语义成分  $\{w_1, w_2, \dots, w_n\}$ ,  $c$  是心理常识复合概念,当且仅当对基础概念集合  $X = \{x_1, x_2, \dots, x_n\}$ , 常识属性集合  $Attr = \{A_1, A_2, \dots, A_p\}$ , 关键字  $S = \{s_1, s_2, \dots, s_q\}$  以及一个子类生成规则  $f$ , 有  $c = f(w_1, w_2, \dots, w_n)$ , 且每个  $w_i$  是: (1)  $X$  中心理常识基础概念  $x_i$  或其子概念, 或; (2)  $Attr$  中某个心理常识属性  $A_i$  的取值  $v \in D_{A_i}$ , 或; (3)  $S$  中某个关键字本身  $s_i$ 。

**定义8(心理常识概念间的父子关系)** 假设  $c_1 = \{w_{11}, w_{12}, \dots, w_{1n}\}$ ,  $c_2 = \{w_{21}, w_{22}, \dots, w_{2m}\}$  是两个心理常识概念,  $c_1$  是  $c_2$  的父概念(记做  $c_2 \angle c_1$ ), 当且仅当对  $c_2$  的任一语义成分  $w_{2j}$ , 在  $c_1$  中存在语义成分  $w_{1i}$ , 使得: (1)  $w_{1i} = w_{2j}$ , 或者; (2) 如果  $w_{1i}$  和  $w_{2j}$  是心理常识基础概念, 且  $w_{1i}$  是  $w_{2j}$  的父概念; (3) 如果  $w_{1i}$  和  $w_{2j}$  属于某个心理常识属性的值域, 且  $w_{2j}$  蕴涵  $w_{1i}$ 。

**定义9(心理常识概念间的直接父子关系)** 假设  $c_1 = \{w_{11}, w_{12}, \dots, w_{1n}\}$ ,  $c_2 = \{w_{21}, w_{22}, \dots, w_{2m}\}$  是两个心理常识概念,  $c_1$  是  $c_2$  的直接父概念(记做  $c_2 | \angle c_1$ ), 当且仅当除  $c_1$  中语义成分  $w_{1i}$  和  $c_2$  的语义成分  $w_{2j}$  外, 其他各语义成分都相等, 且 (1) 若  $w_{1i}$  和  $w_{2j}$  是心理常识基础概念, 则  $w_{1i}$  是  $w_{2j}$  的直接父概念,  $w_{1i}$  可为空概念; (2) 若  $w_{1i}$  和  $w_{2j}$  属于某心理常识属性的值域, 且  $w_{2j}$  直接蕴涵  $w_{1i}$ ,  $w_{1i}$  可为空属性值。

**定义10(心理常识概念的叶子结点)** 设  $c_1 = \{w_{11}, w_{12}, \dots, w_{1n}\}$  是心理常识概念,  $c_1$  是心理常识叶子结点, 当且仅当不存在任一  $c_2 = \{w_{21}, w_{22}, \dots, w_{2m}\}$ , 使得  $c_1$  是  $c_2$  的父概念。

**公理1** 心理常识概念之间的父子关系  $\angle c$  满足: (1) 自反性:  $c_1 \angle c_1$  成立; (2) 反对称性: 如果  $c_2 \angle c_1$  成立, 则  $c_1 \angle c_2$  一定不成立; (3) 传递性: 如果  $c_1 \angle c_2$  成立, 则  $c_2 \angle c_3$  成立, 则  $c_1 \angle c_3$  成立。

下面给出自动生成心理常识复合概念的算法1。

**算法1 心理常识复合概念的生成算法**

- 输入:** 基础概念集  $X = \{x_1, x_2, \dots, x_n\}$ , 常识属性  $Attr = \{A_1, A_2, \dots, A_p\}$  和子类生成规则  $f$ ;
- 输出:** 心理常识概念框架表示及其层次结构;
- Step1:** 读入  $f$  中的语义成分, 包括基础概念集合  $X = \{x_1, x_2, \dots, x_n\}$ , 心理常识属性  $Attr = \{A_1, A_2, \dots, A_p\}$ , 令  $i := n + p$ ;
- Step2:** 分别选取  $f$  的语义成分中可选的  $i$  个, 设为  $\{s_1, s_2, \dots, s_i\} \subseteq \{x_1, x_2, \dots, x_n\} \cup \{A_1, A_2, \dots, A_p\}$ ;
- Step3:** 如果  $s_i \in \{x_1, x_2, \dots, x_n\}$ , 深度优先遍历  $s_i$  及其所有下位概念; 如果  $s_i \in \{A_1, A_2, \dots, A_p\}$ , 深度优先遍历所有属性值以及每一属性值蕴涵的属性值; 设  $s_1 \times s_2 \times \dots \times s_i$  为  $i$  个各语义成分的笛卡尔集  $C$ , 设  $C = \{t_1, t_2, \dots, t_n\}$ , 令  $j := 1$ ;
- Step4:** 选  $C$  中  $t_j = \{w_{1j}, w_{2j}, \dots, w_{ij}\}$ , 依据  $f$ , 组合  $\{w_{1j}, w_{2j}, \dots, w_{ij}\}$  和  $f$  的不可选项, 生成心理常识复合概念集合  $c$ ;
- Step5:** 初始化心理常识复合概念  $c$  的框架; 依次去掉  $\{w_{1j}, w_{2j}, \dots, w_{ij}\}$  中一个语义成分, 将得到的概念名依次放入“是子类:”中;
- Step6:** 对每一个  $\{w_{1j}, w_{2j}, \dots, w_{ij}\}$ , 判断每一个  $w_{ij}$ , 如果  $w_{ij}$  是基础概念, 按  $f$  中  $replace$  项的约定, 将相应槽添加到  $c$  的框架中, 判断  $w_{ij}$  有无直接父概念, 如有则替换现有部分, 将新概念添加到  $c$  框架的“是子类:”中; 如果  $w_{ij}$  是常识属性取值, 按  $f$  中  $replace$  项的约定, 将相应槽添加到  $c$  的框架中, 判断  $w_{ij}$  有无直接蕴含某属性值, 如有则替换现有部分, 将新概念添加到  $c$  框架的“是子类:”中;
- Step7:**  $j := j + 1$ ; 如果  $j < n + 1$ , 则转 Step4;
- Step8:**  $i := i - 1$ ; 如果  $i < 1$ , 则停止, 否则, 转 Step2;

例5 图8是算法1执行的示例。

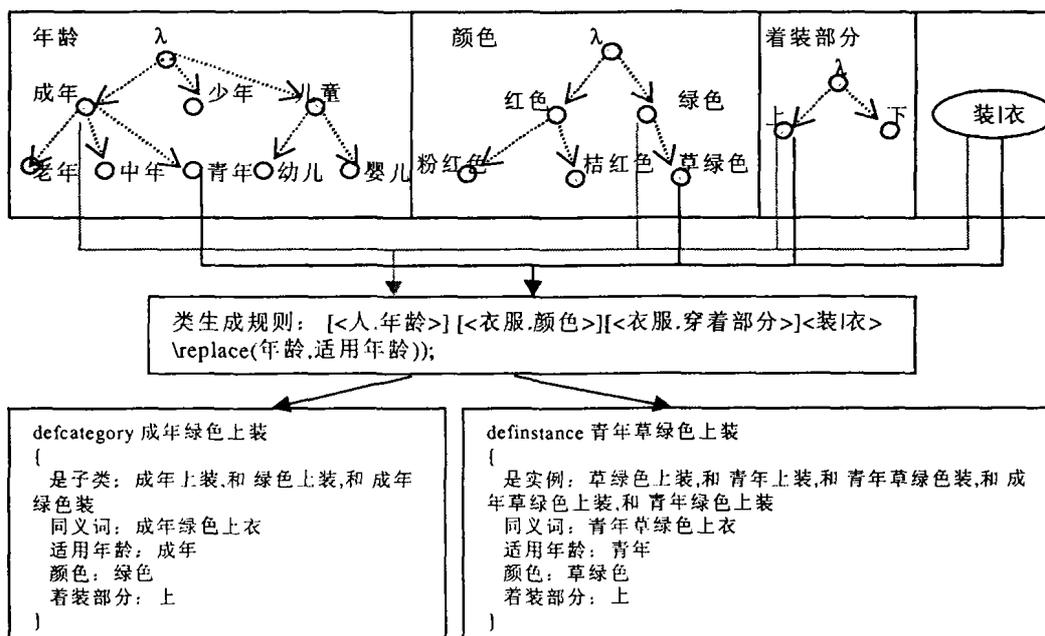


图8 心理常识复合概念的生成示例

上例中,由于属性值“成年”和“青年”间存在蕴涵关系,使得概念“青年草绿色上装”是概念“成年草绿色上装”的直接子

概念,该关系在“成年草绿色上装”中用“有子类”表示,通过该算法,“成年草绿色上装”是“成年绿色上装”的直接子概念,由父子概念关系的传递性,“青年草绿色上装”是“成年绿色上装”的间接子概念。“子类生成规则”中用“〈〉”括起各语义成分,“[ ]”表示该语义成分为可选项;“人. 年龄”表示该语义成分来自于“人”的“年龄”属性的取值。“〈装|衣〉”表示“装”和“衣”两个关键字为不可兼或的选项,且分别以“装”和“衣”生成的复合概念为同义词。

由此,通过心理常识基础概念的层次结构,以及心理常识属性的蕴涵关系,所有心理常识概念以“ $\angle_c$ ”为基础组织成为一个巨大的心理常识概念体系。但我们看到,如果算法 1 中,假设子类生成规则  $f$  的语义成分数为  $m$ ,每个语义成分中的基础概念数(或心理常识属性取值)个数分别为  $\{x_1, x_2, \dots, x_m\}$ ,由于子类生成规则中某些项是可选项,整个算法的复杂性为  $2^{m+1} \times x_1 \times x_2 \times \dots \times x_m$ ,以指数幂的速度增长,随  $m$  增大,执行效率急速下降。另外,如果子类生成规则中本身有子类生成规则,则效率更低。所以,在书写子类生成规则时需控制其规模,我们的策略是:

- (1)控制在“子类生成规则”中使用过多的语义成分,选择语义成分的个数为 1~5;
- (2)控制在“子类生成规则”中使用到上下文相关的各种语义成分;
- (3)控制使用本身的框架中包括了“子类生成规则”的概念作为语义成分。

上述控制限制了生成的心理常识复合概念的语义成分的个数和长度,这样的处理并非全无道理。实际上,由于子类生成的复杂性,以及自然语言中概念的丰富性,完全通过“子类生成规则”生成所有概念也不可行。而且太长的心理常识概念往往意义不大,例如“中国大号红色手工制作的女性保暖毛线外套”。所以我们控制“子类生成规则”的规模,并在新心理常识概念加入时,通过“子类检查规则”来检查它和现有概念库的一致性和冗余性。需要强调,“子类生成规则”和“子类检查规则”在语义和语法上均一致,“子类生成规则”是“子类检查规则”的简化形式,用于生成一些较简单的心理常识复合概念,“子类检查规则”用于检查新加入的常识概念,以下将详细介绍。

#### 4 常识概念的冗余性和一致性分析

在自然语言理解,知识库的建设等过程中,往往需要在常识概念库中追加常识概念。知识工程师对概念的不同理解和用词习惯导致常识概念间产生冗余或不一致,所以需要引入对新加入的概念的检查机制。我们假设待加入的概念已经事先正确分开了各语义成分,下面详细分析新加入概念和现有概念库的冗余和不一致性。一般地,概念的冗余性表现在“异名同义”上,冗余由以下几种情况产生:

- (1)基础概念的同义词;
- (2)心理常识属性的属性值的同义词;
- (3)修饰成分的顺序不一致,如“小号红色服装”和“红色小号服装”;
- (4)修饰成分之间的依赖关系(上下文相关性)导致复合概念的“义”同但“型”不同,如“中国傣族食品”和“傣族食品”。本身“傣族”就只属于中国,所以“中国”作为语义成分来说,是多余的成分。

以上 4 种冗余都可以通过“子类检查规则”来进行纠正。

(1)通过基础概念框架中的“同义词”槽可检查到第一种冗余性。如“衣服”框架中指明其“同义词”是“服装”,则新加入

的“中国衣服”与“中国服装”同义,无需创建两个独立的框架;

(2)心理常识属性值本身是心理常识概念,与(1)相同,其框架中的“同义词”槽可检查出第二种冗余性。例如,“颜色”的属性值包括“红色”,“红色”的框架中“同义词”的值为“赤色”,所以,“红色衣服”和“赤色衣服”是同义词;

(3)〈子类检查规则〉的文法可表示两个语义成分之间可以互换位置而不影响概念的语义的关系,由此可检查出顺序不同但语义相同的复合概念。

(4)第四种冗余性较难于察觉,由于在书写“子类检查规则”时,已控制上下文相关的语义成分同时出现,所以可部分地控制由此产生的冗余性。

新加入的常识复合概念与现有概念库的不一致性表现在以下几种形式:

- (1)各语义成分搭配不当,例如“食味”和关键字“衣服”搭配则不当;
- (2)各个语义成分的取值不相匹配,如“国家”和“民族”两个语义成分间存在上下文相关性,如果取值为“美国”和“傣族”,则不相匹配;
- (3)常识概念的语义成分的顺序不正确,例如“挡风中国衣服”和“傣族中国服装”等;

以上的 3 种冗余都可以通过“子类检查规则”来进行纠正:

- (1)通过“子类检查规则”中选择的可搭配的语义成分,可控制第一种不一致性的产生;
- (2)在书写“子类检查规则”时,控制使用上下文相关的语义成分,可部分控制第二种不一致的产生;
- (3)通过建超“子类检查规则”中语义成分间可互换和不可互换的说明,可检查出第三种不一致性。

子类检查规则的实现同样需要查找新加入的常识概念的语义成分所在的基础概念层次结构或者常识属性,所以也面临执行效率的问题,为加快子类检查算法的效率,我们事先对常识概念库中的基础概念和心理常识属性的取值进行全文索引的二维链表,并在每一项目后标记其出现在哪些基础概念的层次结构(或哪些心理常识属性的值域)中,以及其所在的位置,以方便查找并加快查找速度,索引链表的示例如图 9 所示。

例 6 图 9 是心理常识基础概念和心理常识属性的属性值域的索引链表表示例。

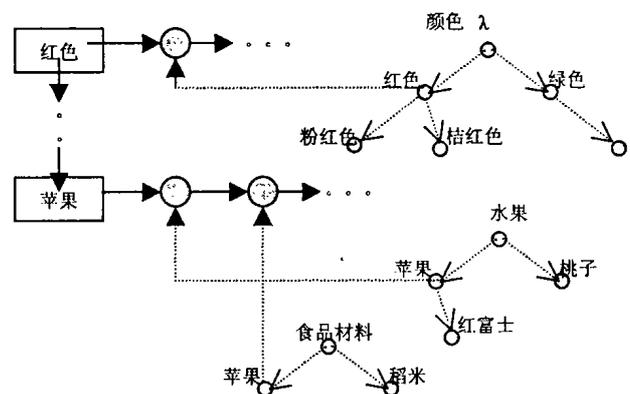


图 9 心理常识基础概念和属性的索引

由于心理常识基础概念可作为独立的概念出现,也可能作为心理常识属性的取值出现,所以其后的链接可能有不止一个。算法 2 是复合常识概念的检查算法。图 9 的结构在检查程序运行之前一次性建立并读入内存中。

#### 算法 2 常识概念的检查算法

(下转第 43 页)

- Distributed Systems [A]. In: Proc. of the Intl. Teletraffic Congress'99(ITC-16), Edinburgh, 1999
- 2 Maria K. Overload control strategies for distributed communication networks; [Ph. D. dissertation]. Sweden: Lund Institute of Technology, 1999
  - 3 王田. SMSC 负荷检测方法研究. 电子科技大学学报[J], 2003, 2(4)
  - 4 Paxson V. Growth trends in wide-area TCP connections [J]. IEEE Network, 1994, 8(4): 8~17
  - 5 Crovella M, Bestavros A. Self-similarity in world wide web traffic evidence and possible causes [A]. In: SIGMETRICS'96, Marriott, PA, 1996
  - 6 Cunha C, Bestavros A, Crovella M. Characteristics of WWW client-based traces [D]. Department of Computer Science of Boston University: [TechRep: TR-95-010]. 1995

- 7 Arlitt M, Williamson C. Web server workload characterization: The search for invariants (extended version) [A]. In: ACM SIGMETRICS'96, Marriott, PA, 1996
- 8 Abdulla G, Fox E, Arbams M. Shared user behavior on the world wide web. In: ProcWebNet97, Association for the Advancement of Computing in Education (AACE), 1997. <http://www.aace.org/>
- 9 Standard Performance Evaluation Corporation. SPECweb99 Benchmark User's Guide, 1999. <http://www.spec.org/>
- 10 欧灿辉, 李晓明. Web 服务器性能评测[J]. 计算机研究与发展, 2002, 39(5): 540~547
- 11 Ranganathan M, Acharya A, Saltz J. Distributed resource monitors for mobile objects [A]. In: Proc. of the Fifth Intl. Workshop on Object Orientation in Operating Systems, Seattle, WA, 1996: 19~23

(上接第12页)

**输入:**待加入常识概念  $c = \{w_1, w_2, \dots, w_k\}$ , 心理常识概念库  $B = \{Attr, X, F, S\}$ , 其中,  $Attr = \{A_1, A_2, \dots, A_n\}$  是心理常识属性集合,  $X = \{x_1, x_2, \dots, x_n\}$  是心理常识基础概念集合, 子类生成规则集合  $F = \{F_1, F_2, \dots, F_p\}$ , 子类检查规则集合  $S = \{S_1, S_2, \dots, S_q\}$ ;

**输出:**对  $c$  与  $B$  的冗余性和不一致性的判定;

- Step1: 读入  $c$  中的各个语义成分  $\{w_1, w_2, \dots, w_k\}$ ; 在索引表中查找各语义成分对应的层次结构, 设为  $\{t_1, t_2, \dots, t_k\}$ ;
- Step2: 查找  $S$  中子类检查规则  $S_1$  是  $S$  的子集, 使得  $S_1$  包含所有的  $\{t_1, t_2, \dots, t_k\}$ ;
- Step3: 逐个检查  $c = \{w_1, w_2, \dots, w_k\}$  中各个语义成分是否严格遵守  $S_1$  中子类检查规则;
- Step4: 若存在  $S_2 \subseteq S_1$ , 使得  $c = \{w_1, w_2, \dots, w_k\}$  中各语义成分严格满足  $S_2$  中任一子类检查规则, 如果  $S_2$  不为空, 则  $c$  与心理常识概念库  $B$  一致; 否则返回“不一致”;
- Step5: 检查  $F$  中的子类生成规则集合  $F_1$  是  $F$  的子集, 使得  $F_1$  包含所有的  $\{t_1, t_2, \dots, t_k\}$ , 逐个检查  $c = \{w_1, w_2, \dots, w_k\}$  中各语义成分是否严格  $F_1$  中子类生成规则;
- Step6: 若存在  $F_2 \subseteq F_1$ , 使  $c = \{w_1, w_2, \dots, w_k\}$  中各语义成分严格满足  $F_2$  中任一子类生成规则, 如果  $F_2$  为空, 则根据语义成分的值, 添加该概念到概念库合适位置, 停止算法; 否则, 返回“有冗余”;

算法2的时间复杂性远低于算法1. 由于在初始化之前建立了心理常识基础概念的索引, 设所有心理常识基础概念的数量是  $n$ , 则算法4.1按折半查找的复杂性是  $n \log_2 n$ . 由于心理常识基础概念对应的结构或常识属性已在每一心理常识基础概念后给出了链接, 因而查找的时间复杂性为常量级. 找到  $F_1$  后检查待检查概念的语义成分与  $F_1$  中(子类检查规则)的匹配性的时间复杂性为语义成分的数量  $m$  (通常  $m$  小于30) 和(子类生成规则)数量  $k$  (子类生成规则的数量也为有限值) 的乘积. 则算法2的复杂性为  $k \times m \times n \times \log_2 n$ . 由于  $m$  和  $k$  的数量都是常数级别的, 所以, 算法2的时间复杂性为  $|n \log_2 n|$ .

**结束语** 本文讨论了与人的心理相关的常识概念的获取、组织和分析. 目的在于克服现有概念模型中缺乏自动获取概念的方法, 并且以心理学范畴为基础, 在手工获取心理常识基础概念和心理常识属性的基础上, 给出自动生成心理常识复合概念以及组织它们的算法, 进一步, 对于新加入概念, 提供算法考察其与现有概念库的冗余问题和一致性问题, 本文的方法为以心理常识概念为基础的心理常识库的建构提供了良好的概念基础.

同时, 由于心理常识复合概念的生成依赖于(子类生成规则), 所以如何正确书写、组织和规范(子类生成规则)是心理常识概念获取的关键, 本文提供了一些基准原则, 但是, 自然语言非常丰富, 我们仍然需要对复合概念的语义成分的构词进一步分析, 并在此基础上加强与语言学家的合作, 构造出更为精确的构造复合概念的规则.

### 参考文献

- 1 Guha R V, Lenat D. Cyc: A midterm Report. AI magazine, 1990. 33~59
- 2 Lenat D. The Dimensions of Context-Space. <http://www.cyc.com>, Oct. 1998
- 3 Churchland P M. Folk Psychology and the Explanation of Human Behavior. In: Churchland PM, ed. A Neurocomputational Perspective, MIT Press, 1989. 111~127
- 4 Stich S, Ravencroft I. What is Folk Psychology?. in Cognition, 1994. 50: 447~468
- 5 Ravenscroft I. Folk Psychology as a Theory, in Stanford Encyclopaedia of Philosophy, 1997
- 6 CYCorp, Feature of CycL. <http://www.cyc.com>, 1998
- 7 陆汝钤, 姬广峰. 基于常识的研究, 世纪之交的知识工程与知识科学. 清华大学出版社, 2001. 469~505
- 8 Richardson S D, Dolan W B, Vanderwende L. MindNet: Acquiring and Structuring Semantic Information from Text. In: Proc. of COLING '98, 1998
- 9 Vanderwende L. Ambiguity in the acquisition of lexical information. In: Proceedings of the AAAI 1995 Spring Symposium Series, working notes of the symposium on representation and acquisition of lexical knowledge, 1995. 174~179
- 10 Fellbaum C. WordNet: an electronic lexical database. The MIT Press, 1998
- 11 Baker C F, Fillmore C J, Lowe J B. The Berkeley FrameNet Project. In: Proc. of the COLING-ACL, Montreal, Canada, 1998
- 12 梅家驹, 竺一鸣, 等. 同义词词林. 上海辞书出版社, 1983
- 13 董振东. 语义关系的表达和知识系统的建造. 语言文字应用, 1998(3): 76~82
- 14 董振东, 董强. 知网介绍. <http://www.keenage.com>, 1999
- 15 Osherson D N, Smith E E. On the Adequacy of Prototype Theory as a Theory of Concepts. in Cognition, 1981, 9: 35~58
- 16 Cohen B. Models of Concept. Cognitive Science, 1984, 8: 27~58
- 17 Palmer S E. Fundamental Aspects of Cognitive Representation. In: E. Rosch, B. L. Lloyd, ed. Cognition and categorization, 1978. 259~302
- 18 薄冰. 高级英语语法. 世界知识出版社, 2000
- 19 Correl S F, Porter B W. Iterative Refinement of Knowledge Bases with Consistency Guarantees, AAAI 1997 Spring Symposium on Ontological Engineering, 1997