

基于血缘关系结构的亲属关系推理算法研究

卢达威

中国人民大学文学院/中国人民大学数字人文研究院 中国人民大学文学院

北京市海淀区中关村大街59号

wedalu@163.com

杨思琴*

北京市海淀区中关村大街59号

yang-sq19@mails.tsinghua.edu.cn

摘要

以往的亲属关系推理系统，对推理的正确性无法保证，对复杂的亲属关系推理容易出错；而且难以解决多个亲属关系作为已知条件的亲属关系推理问题。本文在卢达威等（2019）的基础上，首先将推理规则和推理过程形式化和算法化；进而与基于一阶谓词逻辑的推理系统进行了对比，发现基于血缘关系结构的亲属关系推理在知识表示方法和推理规则方面都存在优势，主要表现在于执行效率更高，以及在编写和核查规则时更不容易出错；最后讨论了亲属关系推理算法的时间复杂度问题，发现该推理系统为是线性时间复杂度。本文的算法及其有效性分析得到了实验结果的支持。

关键词： 血缘关系结构；亲属关系推理；知识表示；推理系统

A Study on Kinship Inference Algorithm Based on Blood Relationship Structure

LU Dawei

School of Liberal Arts, RUC /

School of Digital Humanities of RUC

wedalu@163.com

YANG Siqin*

School of Liberal Arts, RUC

yang-sq19@mails.tsinghua.edu.cn

Abstract

The correctness and completeness of previous kinship inference systems are not guaranteed, which leads to error-prone inferences when the kinship is complex. Moreover, these systems are difficult to infer with multiple kinships as known conditions. Based on Lu et al. (2019), this paper first formalizes and algorithmizes the inference rules and processes. When compared with kinship inference systems based on first-order predicate logic, it is found that the system based on blood relationship structure has advantages in both knowledge representation methods and reasoning rules, particularly in the high efficiency and accuracy of writing and checking rules. Finally, the time complexity problem of the kinship inference algorithm is discussed, revealing that the time complexity of this system is linear. These studies contribute to the practicality of kinship inference.

Keywords: blood relationship structure, kinship inference, knowledge representation, inference system

本文得到国家社会科学基金青年项目“汉语话题延续与转换机制及其计算模型研究”(18CYY030)的资助，特此谢

1 引言

亲属关系推理 (kinship inference) 是知识推理的经典问题。也因其经典性, 容易让人忽略其复杂性。表面上亲属关系推理是语言中亲属称谓词之间的词汇概念的推理。但是, 常见的词汇关系刻画手段, 例如同义—反义关系 (synonymy-antonymy), 上位—下位关系 (hyperonym-hyponym), 整体—部分关系 (holonym-meronym) 等, 都不能反映亲属称谓词之间的实质内涵 (卢达威等, 2019)。乔姆斯基 (2015: 13) 认为, 亲属体系具有很多数学系统的性质; 理解亲属关系还需要用上人类解决形式问题的算术能力。人类学家们虽然早就注意到亲属关系的复杂性 (如邓巴 (2016)、哈维 (2006) 等), 但是就推理而言, 许多语言的亲属称谓系统比较简单, 造成对亲属关系推理的困难习焉不察。对汉语来说, 由于亲属称谓系统相对复杂, 反而很好地揭示了亲属关系推理的复杂性。特别是, 当文本中涉及多个亲属之间的关系时, 推理起来更加困难。例如:

例 (1) 已知: 小明是我的表哥, 老张是我的舅舅。问: 老张和小明是什么关系?

这种亲属关系推理涉及多个句子、多个人物, 是文本中亲属关系推理的常见形式。从亲属称谓的定义看, “舅舅”有两种可能, “母亲的哥哥或弟弟”。“表哥”的可能性则更多, 可以是“母亲的兄、弟、姐、妹的年纪比我大的儿子”也可以是“父亲的姐、妹的年纪比我大的儿子”。例 (1) 的推理是比较复杂的, 答案是“老张可能是小明的父亲、伯父、叔叔、舅舅或者舅母的兄弟。”推理时, 除了考虑生育关系、婚姻关系这两种基本的血缘关系外, 还涉及性别、长幼等属性等。

2022年11月, OpenAI公司推出了对话式通用人工智能工具ChatGPT, 对自然语言处理领域产生了广泛的影响。作为一种生成式人工智能 (AI generated content, AIGC) 应用, ChatGPT在语言理解、语言生成、知识及常识推理等方面的能力都展现出了卓越的类人表现。但也有缺点, 其中之一就是生成信息真实性、准确性不足, 以及类似于例 (1) 这样复杂的亲属关系推理, 还无法准确处理 (如图1)。



Figure 1: ChatGPT对例 (1) 的两次回答

2023年4月, 国家网信办发表了《生成式人工智能服务管理办法 (征求意见稿)》, 对生成式人工智能服务提出了一系列规范要求和指导意见, 其中包括要求生成的内容应该真实准确。“真实”要求防止生成虚假信息, “准确”则包括信息推理不出错。确保生成内容真实准确可以有两种途径: 一是从模型上改进, 让模型尊重事实、正确推理; 二是在必要的时候, 对生成的内容包括事实和推理进行二次核查, 此时可采取外接核查系统或人工干预的方法。面对亲属关系推理这种推理复杂但问题界定清晰的情况, 在必要时外接一个专用的微型的推理系统不失为一个可行的选择。

忧。

©2023 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

学界中，已有不少多个条件下的亲属关系推理系统研究，如靳小龙等（2001）、王树西等（2003）、陈振宇等（2009, 2010a, 2010b）等。卢达威等（2019）从亲属关系知识表达和推理规则两方面梳理了已有研究，指出现有系统在亲属关系的知识表示方面不能反映亲属关系实质，在推理规则方面缺乏正确性保证，一旦遇到关系较远的推理，或遇到例（1）这种比较复杂的推理，很容易出错。该文基于社会学研究，区分了血缘关系和亲属称谓词两个推理层次，认为以往亲属关系推理都是基于亲属称谓词的推理，而推理的实质应该是血缘关系的推理；推理系统应该先将亲属称谓词映射为血缘关系，基于血缘关系推理完成后，再映射为亲属称谓词。

但是，卢达威等（2019）存在一定的不足：在知识表示方面，该文虽然重新设计了基于血缘关系的知识表示方法，但是没有对这种知识表示的合理性作充分的说明。在推理算法方面，该文只有推理的元规则及推理过程示意图，并未对推理算法展开形式化描写，也缺乏论证和推导，其正确性也未能予以证明。另外，无论是对已有系统还是对该文提出的方法，都缺乏对推理过程的时间复杂度等的效率分析。

本文拟在卢达威等（2019）的基础上，首先将推理规则和推理过程形式化和算法化；进而与基于一阶谓词逻辑的推理系统进行对比，讨论它们与基于血缘关系结构的亲属关系推理的差异；最后讨论亲属关系推理算法的时间复杂度问题；并开展相应的实验验证。

2 亲属关系推理过程和推理规则的形式表达

本节以“小明是我的表哥，老张是我的舅舅，问：老张是小明的什么亲戚？”为例，探讨亲属关系的推理规则问题。其中推理规则分为元规则和具体规则。元规则指导整个推理过程，具体规则配合元规则实现具体的推理任务。这是使用传统的符号系统的方法来构建的系统。区别于基于统计或生成式的模型，对于小型符号系统，我们要求逻辑分析无误，做到100%的推理准确率。

2.1 基础概念

为更好说明亲属关系推理的形式表达，我们将卢达威等（2019）所定义的血缘关系结构和亲属关系式作为基础概念。

1. 血缘关系结构

血缘关系结构是通过婚姻关系、生育关系为纽带，以“父亲”“母亲”“孩子”组成的小家庭为基础单元的递归结构。血缘关系结构中，为方便推理，定义了四种基本关系：婚姻关系、生育关系、被生育关系、兄弟姐妹关系。通过四种“基本血缘关系”，构建出血缘关系结构图（如Figure2）。血缘关系结构图是推理的基础。

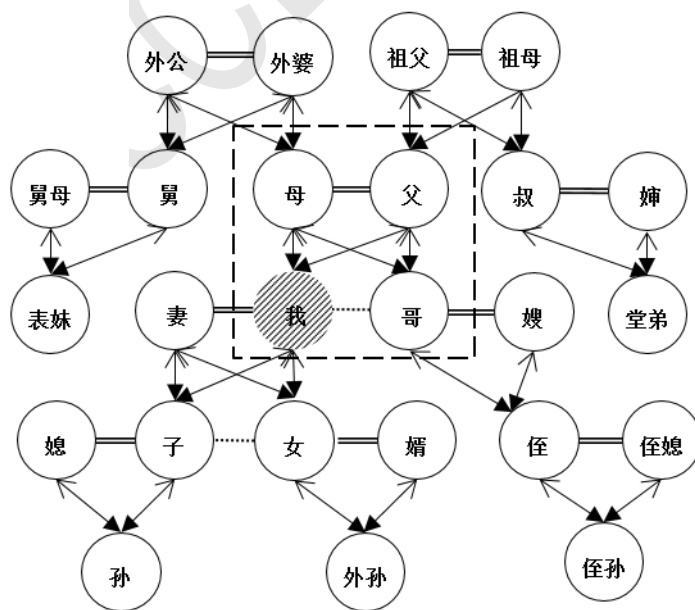


Figure 2: 血缘关系结构图（局部）

2. 亲属关系式

卢达威等 (2019) 没有使用传统的一阶谓词逻辑来表示亲属关系，而是重新构造了亲属关系的知识表示，称为“亲属关系式”，如下：

$$R = s(x_0) + \sum_{i=1}^n f(x_i, y_i)$$

R为亲属关系式，其中， $s(x_0)$ 是起始标记，代表自己， x_0 是性别属性； $f(x,y)$ 代表四种“基本血缘关系”之一，这四种“基本血缘关系”包括婚姻关系 $m(x)$ 、被生育关系 $p(x)$ 、生育关系 $b(x,y)$ 和兄弟姐妹关系 $c(x,y)$ ；参数 x 是性别属性， $x \in \{1, -1, 0\}$ ，1为“男”、-1为“女”、0为“两可/未知”；参数 y 是长幼属性， $y \in \{1, -1, 0\}$ ，1为“长”、-1为“幼”、0为“两可/未知”。该公式的含义是，任何一个亲属关系均可以表达为：从自身出发的若干种基本血缘关系的排列。

2.2 亲属关系推理的元规则

基于此亲属关系知识表示和血缘关系结构图，亲属关系推理元规则可以分为四步：

(1) 查找定义。利用亲属关系定义，在血缘关系图上查找已知条件的亲属表达式。如上例，以“我”为中心在血缘关系图上找到“小明”结点（我的表哥）和“老张”结点（我的舅舅）。其中“表哥”路径有多个，应该分多种情况计算。

(2) 连接。将血缘关系图中所求的两结点连接起来。亲属关系式有向的，例如，所求以“小明”为出发点，则应首先要调用逆关系运算规则，将“我→小明”的有向连接（即“小明是我的...亲戚”），变成“小明→我”的有向连接（即“我是小明的...亲戚”）；然后，以“我”结点作为桥梁，形成“小明→我→老张”的有向连接，得到一条从“小明”到“老张”的亲属关系式（即“老张是小明的...亲戚”）。

(3) 约简。寻找所求两结点的最短路径。通过执行亲属关系式的约简规则，删除路径中不必要的关系，求出“小明→老张”的最短路径。

(4) 匹配。根据最短路径上的基本血缘关系序列，匹配亲属关系定义，得到“老张”对“小明”的亲属称谓。

在以上亲属关系推理元规则中，主要涉及两类规则：一类是步骤（2）中亲属关系的逆关系运算规则；另一类是步骤（3）中寻找最短路径的约简规则。以下分别展开说明。

2.3 亲属关系的逆关系运算规则

亲属关系是有向的。血缘关系结构图中的两个结点A和B，从A到B所经过的血缘关系，与从B到A所经过的血缘关系是不同的。主要体现在三个方面：亲属关系式中基本血缘关系序列排列相反；生育关系 b 和被生育关系 p 不对称；基本关系中属性参数的含义不同。因此，我们要对逆关系运算专门定义。设原亲属关系式R为： $s(0)+f(x_1)+g(x_2)+h(x_3)$ ，其中 f, g, h 是四种基本血缘关系之一，如Figure3，逆关系运算分三步：

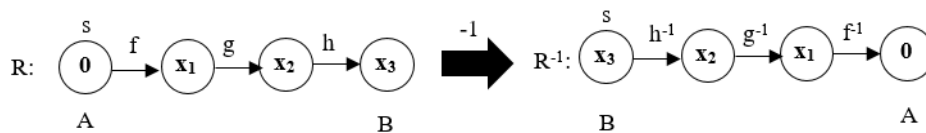


Figure 3: 亲属关系式R的逆关系运算

(1) 逆关系 R^{-1} 中，将原亲属关系式R的基本血缘关系的排列 $f+g+h$ 倒置，即 $h+g+f$ 。

(2) 对每个基本血缘关系分别求逆：婚姻关系 m 和兄弟姐妹关系 c 的逆关系是自身，生育关系和被生育关系互为逆关系。即 $m^{-1}=m$ ， $c^{-1}=c$ ， $b^{-1}=p$ ， $p^{-1}=b$ 。

(3) 重新安排各基本血缘关系的属性。性别属性取值为基本血缘关系的终点结点的性别属性值，即本例中，将求逆前的 $f(x_1)$ ， $g(x_2)$ ， $h(x_3)$ 分别变成 $f^{-1}(x_0)$ ， $g^{-1}(x_1)$ ， $h^{-1}(x_2)$ 。对于长幼属性，则进行属性值取反，原值为0时不变。

经过以上三步，亲属关系式R的逆关系式 R^{-1} 为： $s(x_3)+h^{-1}(x_2)+g^{-1}(x_1)+f^{-1}(x_0)$ 。

2.4 亲属关系式约简规则

2.4.1 约简规则完备性保证

所谓约简，就是在亲属关系式中，用尽可能少的基本血缘关系来表达。亲属关系式的约简，是亲属关系推理中最重要的一步。亲属关系式的约简的完备性，是基于以下两个发现作为保证的：

(1) 关系的约简只可能出现在“父亲”“母亲”“孩子”组成的“血缘关系基础单元”之内。跨越基础单元的两个关系不存在被约简的可能。比如“父亲的父亲 (p+p)”不在同一基础单元内，不能被约简。

(2) 所有约简总能规约为两个关系之间的约简，不存在必须包含3个关系或以上的约简。例如，“兄弟的父亲的妻子 (c+p+m)”，可以通过两两约简，先将“兄弟的父亲 (c+p)”约简成“父亲 (p)”，再将“父亲的妻子 (p+m)”约简成“母亲 (p)”。

以上两个特征说明：只要充分刻画一个血缘关系基础单元内两两关系之间的约简规则，就能够保证整个亲属关系式能够通过这些规则，达到最简。

2.4.2 8种约简规则

对于4种基本血缘关系，两两组合共有16种组合，其中，有8种两两基本关系组合恰好落在同一个血缘关系基础单元中，可被约简成一种基本血缘关系；有8种不在同一个基础单元中，不可被约简。分类如Table1所示。

Table 1: 可被约简的和不可被约简的两两亲属关系组合

可被约简的两两亲属关系	不可被约简的两两亲属关系
m+m	m+p
m+b	m+c
b+p	b+m
b+c	b+b
p+m	p+p
p+b	p+c
c+p	c+m
c+c	c+b

Table1右栏是不能约简的8种基本血缘关系组合，暂时不在我们的讨论范围。左栏是8种可约简的基本血缘关系组合，恰好覆盖了一个血缘关系基础单元内(Figure2) 所有可能的两两关系组合。左栏的两个基本血缘关系均可以约简为一个或0个基本血缘关系，约简规则详列于Table2。(注：在详细约简条件中，“→”号左边的关系和属性是前提条件，右边是约简的结果；f(x)是前一个基本血缘关系，表示约简时会受到前一个基本血缘关系f的性别属性x的影响；w, x, y, z表示属性值，可以是1、-1、0；a, e表示非0的属性值，只能是1、-1；*表示约简时不需要考虑的属性值)

由于约简细节较为复杂，我们仅以公式 (2.3) $b+p \rightarrow 0 / m$ 为例。该公式的字面含义是，“(我的)孩子的父/母 (b+p)”可能是“我自己 (0)”或者“(我的)配偶 (m)”。根据公式中生育关系b、被生育关系p以及前置关系f的性别属性，分为4种情况：

公式 (2.3-1) : $f(a) + b(*, *) + p(a) \rightarrow f(a)$

表示若前置关系f的性别已知为a (a不为0)，父/母亲p的性别也已知为a (即两者相等且不为0)，则“(f的)孩子的父/母”必然是“(f)自己”，性别为a。

公式 (2.3-2) : $f(a) + b(*, *) + p(-a) \rightarrow f(a) + m(-a)$

表示若前置关f的性别已知为a，p的性别也已知为-a (即性别恰好相反)，则“(f的)孩子的父/母”必然是“(f的)配偶”，且性别与p相同。

公式 (2.3-3) : $f(a) + b(*, *) + p(0) \rightarrow f(a) / f(a)+m(-a)$

表示若前置关f的性别已知为a，但p的性别属性值为0 (两可)，则“(f的)孩子的父/母”可能是“(f)自己”，性别为a，也可能是“(f的)配偶”，性别为-a。

公式 (2.3-4) : $f(0) + b(*, *) + p(x) \rightarrow f(x) / f(-x)+m(x)$

表示若前置关f的性别未知，则无论p的性别如何，“(f的)孩子的父/母”都有两种可能：可能

Table 2: 亲属关系约简规则

序号	规则简写	公式含义	公式编号	详细约简条件
1.	$m+b \rightarrow b$	(我的)配偶的孩子 \rightarrow (我的)孩子	(2.1)	$m(*)+b(x,*) \rightarrow b(x,*)$
2.	$m+m \rightarrow 0$	(我的)配偶的配偶 \rightarrow (我)	(2.2)	$m(x)+m(y) \rightarrow 0$ 且 $x=-y$
3.	$b+p \rightarrow 0/m$	(我的)孩子的父/母亲 \rightarrow (我)或(我的)配偶	(2.3)-1 -2 -3 -4	$f(a)+b(*)+p(a) \rightarrow f(a)$ $f(a)+b(*)+p(-a) \rightarrow f(a)+m(-a)$ $f(a)+b(*)+p(0) \rightarrow f(a) / f(a)+m(-a)$ $f(0)+b(*)+p(x) \rightarrow f(x) / f(-x)+m(x)$
4.	$b+c \rightarrow b$	(我的)孩子的兄弟姐妹 \rightarrow (我的)孩子	(2.4)	$b(*)+c(x,y) \rightarrow b(x,k(x,y,z))$ $k(x,y) = \begin{cases} 1, & x=y=1 \\ -1, & x=y=-1 \\ 0, & \text{other} \end{cases}$
5.	$p+b \rightarrow 0/c$	(我的)父/母亲的孩子 \rightarrow (我)或(我的)兄弟姐妹	(2.5)-1 -2 -3 -4	$f(a)+p(*)+b(-a,x) \rightarrow f(a)+c(-a,k(x,y))$ $f(x)+p(*)+b(y,e) \rightarrow f(x)+c(y,e)$ $f(0)+p(*)+b(x,0) \rightarrow f(x) / f(0)+c(x,0)$ $f(a)+p(*)+b(x,0) \rightarrow f(a) / f(a)+c(x,0)$
6.	$p+m \rightarrow p$	(我的)父/母亲的配偶 \rightarrow (我的)父/母亲	(2.6)	$p(x)+m(y) \rightarrow p(y)$ 且 $x=-y$
7.	$c+p \rightarrow p$	(我的)兄弟姐妹的父/母亲 \rightarrow (我的)父/母亲	(2.7)	$c(*)+p(x) \rightarrow p(x)$
8.	$c+c \rightarrow 0/c$	(我的)兄弟姐妹的兄弟姐妹 \rightarrow (我)或(我的)兄弟姐妹	(2.8)-1 -2 -3	$f(a)+c(*)+c(-a,y) \rightarrow f(a)+c(-a,k(x,y))$ $f(w)+c(*)+c(z,a) \rightarrow f(w)+c(z,a)$ $f(w)+c(*)+c(z,y) \rightarrow f(z) / f(w)+c(z,k(x,y))$ $k(x,y) = \begin{cases} 1, & x=y=1 \\ -1, & x=y=-1 \\ 0, & \text{other} \end{cases}$

注：“规则简写”栏中的基本亲属关系与“详细约简条件”栏中的波浪线对应。

是“(f)自己”，此时性别与p一致，为x；或可能是“(f的)配偶”，性别与p一致，并推论出f的性别与p相反，为-x。

以上4个子公式的属性组合以及覆盖了公式 (3.3) $f+b+p$ 所有可能的性别属性组合，没有遗漏。其他的约简公式与此类似。

2.5 亲属关系式的约简流程

对于亲属关系式 $R=s+f_1+f_2+\dots+f_n$ (f代表基本血缘关系) 约简流程大致描述如下。

从左到右逐一扫描亲属关系式R中当前一个及后续一个基本血缘关系 (如 f_i+f_{i+1})，约简情况有三种：

- (1) 若不能约简，则直接执行下一个两两基本血缘关系组合 (如 $f_{i+1}+f_{i+2}$)；
- (2) 若能够约简，则约简并以约简后的结果为起点，扫描后续两两基本血缘关系组合；
- (3) 若执行步骤 (2) 时产生新了亲属关系式分支，则记录下来，完成当前亲属关系式约简后再对新的分支进行约简 (重复步骤1-3)。

以上约简流程和Table2的8个约简规则，有效保证了亲属关系式能最大程度约简，从而求得两个结点之间的最短路径。然而，除了以上两类规则外，推理过程中，还有一些细节需要定义相应规则。如亲属关系式连接规则和属性核查方法等，篇幅所限，不作赘述。

2.6 亲属关系推理过程举例

下面以“已知：小明是我的表哥，老张是我的舅舅，问：老张和小明是什么关系？”为例，说明推理过程和规则的使用方法。根据2.2节推理的元规则 (①查找—②连接—③约简—④匹配) 流程如Figure4。

详细解释如下。

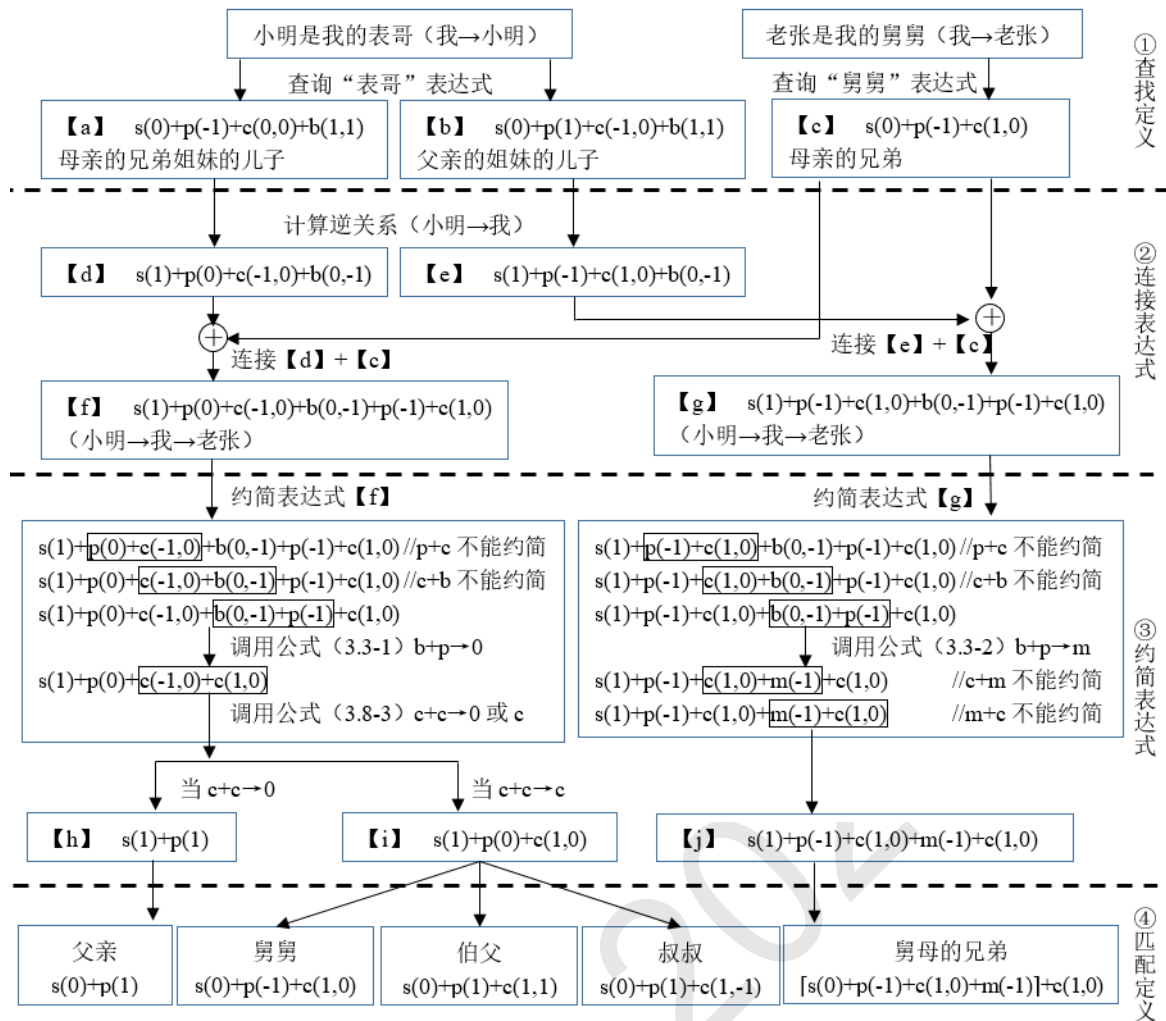


Figure 4: 亲属关系推理流程

(一) 查找定义：表哥和舅舅

查询亲属关系式可知，“表哥”的亲属关系式有两种（Figure4的公式【a】和【b】），“舅舅”的亲属关系式有一种（公式【c】）。

(二) 连接表达式：连接小明和老张

我们以“老张是小明的XXX亲戚”为所求，则“小明”是连接的起点，“老张”是终点。形成“小明→老张”的亲属关系式。“根据题目，“小明”和“老张”的媒介是“我”，所以连接的最终结果是“小明→我→老张”。分为逆关系变换和连接两步。

逆关系变换：已知条件中，关于“表哥”的定义有两个，【a】和【b】，其方向都是“我→小明”的，要首先调用逆关系规则，将【a】和【b】分别变成【d】和【e】，使亲属关系的方向从“我→小明”变成“小明→我”。

连接：利用亲属关系式连接规则将【d】和【e】“小明→我”与【c】“我→老张”连接起来，形成【f】和【g】“小明→我→老张”。

(三) 约简表达式：约简小明到老张的亲属关系式

由于“表哥”的亲属关系定义关系有两种，因此连接后“小明→老张”的亲属关系式也有两个，要分别约简。

对于【f】 $s(1)+p(0)+c(-1,0)+b(0,-1)+p(-1)+c(1,0)$ ，先从左到右两两基本血缘关系开始约简，过程如下：

- (1) 检查 $p(0)+c(-1,0)$ ，查找约简规则发现 $p+c$ 不能被约简。
- (2) 检查下一组两两关系 $c(-1,0)+b(0,-1)$ ，发现 $c+b$ 也不能约简。
- (3) 检查下一组两两关系 $b(0,-1)+p(-1)$ ，发现 $b+p$ 是可以约简的，根据约简规则 (2.8)，

由于p的性别属性值-1与前置关系c的性别属性值相同，应该调用公式（2.8-1），删除 $b(0,-1)$ 和 $p(-1)$ ，意义是“某人（女）孩子的母亲”是“她自己”。

(4) 因为原关系经过了约简删除，重新核查 $c(-1,0)+c(1,0)$ ，发现 $c+c$ 是可以简化的，根据约简规则（2.8），前置关系的性别不确定（值为0）时，应该调用公式（2.8-3），产生两种可能亲属关系式，一种是删除 $c(-1,0)$ 和 $c(1,0)$ ，指“某人的姐妹的兄弟”是“某人自己（男）”；另一种是删除 $c(-1,0)$ ，指“某人的姐妹的兄弟”是“某人（性别不确定）的兄弟”。

(5) 新产生的两条亲属关系式均已达到关系式末尾，也没有新的关系式需要约简。约简结束。

(6) 经过约简，产生了两条约简结果，即： $s(1)+p(1)$ 和 $s(1)+p(0)+c(1,0)$ ，对应公式【h】和【i】。

对于【g】，约简流程相似，只是调用的规则稍有不同，此处不在赘述。约简后的亲属关系式是公式【j】： $s(1)+p(-1)+c(1,0)+m(-1)+c(1,0)$ 。

(四) 匹配亲属关系定义

经过约简，得到“小明→老张”的亲属关系式有三个，【h】 【i】 【j】。

通过匹配亲属关系式的定义可知，【h】是“父亲”，即“老张是小明的父亲”。

而【i】的字面含义是“老张是小明的父亲或母亲的兄弟”，根据亲属关系定义有多种可能。“母亲的兄弟”是“舅舅”，“父亲的哥哥”是“伯父”，“父亲的弟弟”是“叔叔”。

再看【j】。其含义是“母亲的兄弟的妻子的兄弟”，这已经超出了一般亲属关系定义的范围。根据最大匹配原则，“母亲的兄弟的妻子”可以匹配为“舅母”。所以，【j】用两段亲属关系来描述，匹配为“舅母的兄弟”。

由此，我们知道，当“小明是我的表哥”，“老张是我的舅舅”，那么“老张”有可能是“小明”的“父亲”“舅舅”“伯父”“叔叔”或“舅母的兄弟”。

3 基于血缘关系结构的亲属关系推理的优势

3.1 亲属关系知识表示的优势

我们以“舅舅”为例，分析亲属关系式与一阶谓词逻辑的不同。“舅舅”亲属关系式为： $s(0)+p(-1)+c(1,0)$ ，而使用传统的一阶谓词逻辑，可能表示为： $亲子(z1, x) \wedge 亲子(z2, z1) \wedge 亲子(z2, y) \wedge 女(z1) \wedge 男(y)$ 。

首先，一阶谓词逻辑的表达一般需要将经过的结点（ $z1, z2, z3$ ）假设出来，并作为变量。在推理时，需要对这些参数进行合一运算，而合一运算时可能涉及若干回溯，增加了推理的时间复杂度；而亲属关系式则直接使用了基本血缘关系，隐藏了结点，故不需要假设名称，也不需要合一运算。

第二，亲属关系式将属于结点的性别属性、长幼属性等附着到基本血缘关系中，更直观，使得编写和检查亲属关系定义的效率更高，错误率更低。

第三，一阶谓词逻辑一般只用一个特征表示生育关系，如“亲子(x,y)”，而亲属关系式区分了生育关系 $b(x,y)$ 和被生育关系 $p(x)$ ，实际上二者是不对称的。从上节Table2的公式（2.3）和（2.5）的区别可知，“我的孩子的父/母亲”（ $b+p$ ）是“我”或“我的配偶”，而“我的父/母亲的孩子”（ $p+b$ ）是“我的兄弟姐妹”。这一区别，亲属关系式可以通过 b 和 p 的顺序来表达出来。而对于只有“亲子(x,y)”的一阶谓词逻辑，“我的孩子的父/母亲”应表示为“ $亲子(x,z) \wedge 亲子(y,z)$ ”，“我的父/母亲的孩子”表示为“ $亲子(z,x) \wedge 亲子(z,y)$ ”，二者的区别隐含在结点参数的顺序中。这增加了计算和错误检查的难度。当关系复杂或条件众多时，容易引起失误。

第四，亲属关系式能够让所有亲属关系称谓的集合变成一个“可列集”。理论上，只要对四种基本血缘关系及其属性，逐一排列和穷举，就能无遗漏地列出人类所有的亲属关系式。对于推理系统来说，就能够明确亲属关系推理系统的推理范围，构造一个在此范围内的完备的亲属关系知识体系，甚至发现人类知识的盲区。例如，虽然《亲属称呼辞典》中列举了10大类亲属333小类的亲属称谓，但是依然无法确认这是不是亲属关系的全部。如果我们利用亲属关系，就能确定一定的推理范围，逐一列出此范围内的亲属称谓词，就能一一指出在此范围内，有哪些亲属关系是缺乏称谓的，可能会引起对人类学等方面的思考。

3.2 亲属关系推理规则的优势

一般基于一阶谓词逻辑的亲属关系推理，其实质是基于亲属称谓词进行推理的。具体而

言，其推理的过程需要不断与亲属称谓词的定义进对比和进行合一运算，直到找到合适的亲属关系作为答案。这样的做法，虽然大多数情况能计算出正确答案，但一来效率低，时间复杂度高；二来，若亲属关系称谓的定义不完备，很可能因缺乏定义而无法判断当前的逻辑表达式是需要回溯还是推理失败，从而导致推理系统出错。而基于一阶谓词逻辑的知识表达，无法解决亲属关系知识体系的完备性问题，往往靠人的经验来列出亲属关系称谓的定义。当关系较远的时候，人的经验很可能不足，造成定义的不完备。三来，当需要推理的关系复杂的时候，不能保证这类推理的有效性。

也有的系统不是直接匹配，而是先进行逻辑表达式的约简（如陈振宇，2010）。问题是，无法保证提出的约简规则是否覆盖了所有可以约简的情况。这是由于约简目标不明确造成的。

与基于血缘关系结构的推理则不同，基于血缘关系结构的推理系统先把所有亲属关系都映射到血缘关系结构图中，然后在血缘关系结构上进行推理；而且，这样的推理是有明确约简目标的，即找到两个结点的最短路径。推理过程中，我们先论证了所有关系的约简都能归结为两两基本亲属关系的约简；进一步，我们找到了两两关系约简的所有情况，并有针对性设计了约简规则。最后，再把两个结点的最短血缘关系路径映射到亲属称谓词中。

这样做有几个好处：一是，由于有明确的约简目标，能够检验亲属关系的推理结果的准确性；二是效率高，由于推理过程是基于血缘关系结构图的，不需要反复进行比对亲属关系称谓词的合一运算；三是，即使偶有亲属关系称谓词的定义不完备，也不影响推理的过程，因为此时的血缘关系以及最简，无论是否存在亲属称谓的定义，都可以结束推理。不过实际上，由于亲属关系式的可列性，定义的完备是可以保证的。

4 亲属关系推理的时间复杂度分析

对基于血缘关系的亲属关系推理的时间复杂度，我们对元规则的四个步骤（定义、连接、约简、匹配）逐一分析。由于这几个步骤是串行的，推理过程的时间复杂度取决于其中时间复杂度最大的部分。

（一）查找定义的时间复杂度

查找定义是将亲属称谓词转换为亲属关系式，这一过程只需要查表即可，因此是常数时间复杂度的，记为 $O(1)$ 。

（二）连接的时间复杂度

连接操作是将不同的条件，根据起始结点和所求目标，将不同的亲属关系式进行连接。个别亲属关系式需要先进行逆关系操作。逆关系操作和连接操作都是常数时间复杂度的，总体还是常数时间复杂度，记为 $O(1)$ 。

（三）约简的时间复杂度

约简操作是对目标亲属关系式求最短路径。由于我们发现亲属关系式总可以通过两两约简达到最简状态，因此，这一过程是不需要回溯的，而且每一次约简，只需要匹配2.3节的约简规则，进行约简操作，这个过程是常数时间的。就是说，只需要对目标亲属关系式的每个基本血缘关系扫描一遍，即可得到最简亲属关系式。对于 N 个基本血缘关系的亲属关系式，时间复杂度为 $O(N)$ 。

（四）匹配亲属称谓词的时间复杂度

根据最简亲属关系式，我们需要在亲属称谓定义表中找到匹配的亲属称谓。这一过程也是常数时间复杂度的，记为 $O(1)$ 。

综上所述，四个过程的时间复杂度为 $O(1)+O(1)+O(N)+O(1)$ ，则推理系统的时间复杂度是 $O(N)$ ，即线性时间复杂度。 N 是未约简的亲属关系式中基本血缘关系的个数，事实上 N 的规模非常有限。整个推理系统的复杂度是很低的。而且空间复杂度也很低，仅需要存储亲属称谓的定义集和推理规则集，推理过程中使用的空间使用量是一个常数，非常适合于外接到任意系统进行亲属关系的实时推理。

5 亲属关系推理实验

在前文的理论分析的基础上，本节将开展亲属关系推理的实验，对推理的正确性和运行效率进行检验。

首先，我们构造了一个亲属关系称谓集。为了描写亲属关系称谓集，我们定义先“ N 阶亲属关系”的概念。“ N 阶亲属关系”指该称谓的亲属关系式中基本亲属关系数量为 N （如Table 3）。

Table 3: 各阶亲属关系

亲属关系类型	数量	亲属称谓举例	亲属关系式	含义
一阶亲属关系	16	女儿	$s(0)+b(-1, 0)$	女儿
二阶亲属关系	35	祖父	$s(0)+p(1)+p(1)$	父亲的父亲
三阶亲属关系	86	堂哥	$s(0)+p(1)+c(1,0)+b(1,1)$	父亲的兄弟的儿子（比我年纪大）
四阶亲属关系	58	堂妹夫	$s(0)+p(1)+c(1,0)+b(-1,-1)+m(1)$	父亲的兄弟的女儿（比我年纪小）的丈夫

基于N阶亲属关系的概念，亲属关系称谓集中，我们定义了一阶亲属关系称谓16个，包括配偶（丈夫、妻子）、双亲（父亲、母亲）、孩子（儿子、女儿）、兄弟姐妹等；二阶亲属关系称谓35个，如祖父、孙子、舅舅、女婿等；三阶亲属关系称谓86个，如曾祖父、曾孙、表兄、亲家等；四阶亲属关系称谓58个，如堂舅、表嫂、姑表舅父、姨表姨母等。由于四阶亲属关系已经很多称谓不是很熟悉了，所以不再定义五阶亲属关系。值得说明的是，一阶和二阶已经无遗漏地穷举了所有亲属关系表达式。可以穷举、令所有亲属关系变成一个可列集，正是亲属关系式的优势之一。而三阶和四阶则选择了一些典型的为人熟知的亲属关系进行定义。总的来说，本文的亲属关系称谓集共定义了195个亲属关系称谓。

第二，我们将所有关系的两两组合进行全排列，如“爸爸的堂哥”“堂妹夫的表姐”等，则195个称谓共构造出 $195 \times 195 = 38025$ 个亲属关系推理式。

第三，使用本文的算法对这38025个亲属关系式进行推理。对于推理结果，我们随机抽取了10%进行人工验证，即3802个推理结果。这些推理结果包括两类，如果推理后得到的亲属称谓在我们的亲属称谓集中，则直接匹配亲属称谓名；若超过了亲属关系称谓集的范围，则先按最大程度匹配亲属关系称谓，再在此基础上进一步匹配其余部分。例如，“父亲的曾祖母”，由于推理结果是“曾祖父的母亲”。

由于是基于规则的推理，经检验，推理结果达到100%的正确率。这38025个亲属关系推理在普通个人计算机中（i7-4710MQ, 2.5GHz的CPU, 16G内存）运行时间为7秒左右。

可见，本文亲属关系推理的正确性和有效性得到了实验的支持。

6 结语

亲属关系推理是一个经典的常识推理问题，看似简单实质复杂。以往的亲属关系推理系统，或者不能解决文本中多个亲属关系作为已知条件的亲属关系推理问题，或者对复杂的亲属关系推理容易出错，且计算量大，难以实用。

本文基于卢达威等（2019），使用了传统的符号系统的方法来构建亲属关系推理系统。以基于血缘关系的亲属关系式为基础，着重构建了推理规则和推理过程的形式系统；并探讨了本文的基于血缘关系结构的亲属关系推理系统，相对于一阶谓词逻辑的推理系统的优势；还分析了时间复杂度等理论问题。本文虽然使用传统的规则方法，但从理论上能保证100%的推理正确率，并通过专门设计的亲属关系表达方式和推理规则，保证了推理的高效性。相当于在理论上和实践上彻底解决了亲属关系推理的问题。

本文所构造的亲属关系推理系统，是一个封闭、准确、小巧、离线的系统，能够独立推理，也能够非常方便得接入ChatGPT等语言处理系统中，为系统提供精确推理的服务。

值得说明的是，本文的亲属关系推理是基于“血缘关系”这一人类共有的关系构建的，因此是跨语言的。不仅汉语亲属关系推理适用，其他语言只要修改亲属关系定义部分，就可直接进行推理。另外，本文的亲属关系推理，实际上是利用血缘关系构造了一个亲属关系概念的语义空间，该语义空间有四个维度： m, b, p, c ，实现了语义空间内的运算。这为其他语义问题的推理，提供一个借鉴。

参考文献

陈振宇, 袁毓林, 张秀松, 周强. 2009. 亲属关系的逻辑意义及其自动推理. 计算机工程与应用, 45(16): 43-47.

- 陈振宇, 袁毓林, 张秀松, 周强. 2010a. 一种基于大知识库的亲属关系自动推理模型. 中文信息学报, 24(3): 117-124.
- 陈振宇, 袁毓林. 2010b. 汉语亲属关系的语义表示和自动推理. 中国语文, (1): 44-56.
- 邓巴著 (余彬译). 2016. 人类的演化. 上海: 上海文艺出版社.
- 官赛萍, 靳小龙, 贾岩涛, 王元卓, 程学旗. 2018. 面向知识图谱的知识推理研究进展. 软件学报, 29(10): 2966-2994.
- 哈维兰著 (瞿铁鹏, 张钰译). 2006. 文化人类学(第10版). 上海: 上海社科院出版社.
- 靳小龙, 魏旺强. 2001. 基于常识的亲属关系推理模型. 计算机工程与应用, 37(17): 83-85.
- 卢达威, 袁毓林. 2019. 基于血缘关系结构图的亲属关系推理系统研究与实现. 中国社会科学, (11): 25-43.
- 诺姆·乔姆斯基著 (曹道根, 胡鹏志译). 2015. 语言的科学. 北京: 商务印书馆.
- 王树西, 刘群, 白硕. 2003. 一个人物关系问答的专家系统. 广西师范大学学报(自然科学版), (1): 31-36.