计算机科学1993 Vol. 20 No.4

自动定理证明：十年回顾*)

贲可荣　陈火旺

（国防科技大学计算机系　长沙410073）

摘　要

本文主要介绍近十年来定理证明器的发展情况，同时讨论了与自动定理证明相关的一些问题。

推理的自动化（包括自动定理证明）特别诱人，因为所有数学以及许多技术领域均可用一定的形式系统来表述。一阶逻辑是一种特别简单而易于理解的语言，足以表达许多问题，甚至von Neuman-Bernays-Gödel集合论。今天用现代计算机进行快速而准确的推理已成为可能。另一方面，自动推理可以避免令人乏味和容易出错的详细证明构造过程。现在的问题不在速度，而主要在控制策略，计算机以同样的灵活性产生相关的或不相关的结果。

定理证明器的可能应用包括专家系统、规划、证明验证、教育以及作为数学家的工具。事实上，可以想象，哪里用到推理，哪里就用到定理证明器。可以将一定理证明器看做是最终的说明性程序设计语言，或者用来实现一程序设计语言的高层控制。

1. 历史的回顾

1956年，Newell，Shaw和Simon给出了一个称为"逻辑机器"的程序，证明了罗素、怀德海所著《数学原理》中的许多定理，这标志着自动定理证明的开端。

1959年，Gelernter给出了一个称为"几何机器"的程序，能够做一些中学的几何题，速度与学生相当。

1960年，美籍华裔王浩，在IBM704上，编程实现了三个程序。第一个程序用于命题逻辑，第二个程序让机器从基本符号出发自动生成合适命题逻辑公式并选出其中定理，第三个程序用于判定一阶逻辑中的定理。他证明了罗素、怀德海《数学原理》中的几乎所有定理。他的方法，人们称之为"王浩"算法。他的这项工作，在1984年首届自动定理证明大会上获最高奖——"里程碑"奖。

1960年，Davis，Putnan等给出了D-P过程，大大简化了命题逻辑的处理。

1965年，Robinson提出了归结方法，使得自动定理证明领域发生了质的变化。

1968年，苏联Maslov提出了逆向法，是苏联人六、七十年代在该领域做的主要工作。

归结方法有许多重要改进，每种改进有各自的优点。语义归结，由Slagle（1967）提出，它将超归结(Robinson 1965)、换名归结（Meltzer，1966）、支架集策略(Wos，Robinson，1965）等方法一体化。锁归结，由Boyer（1971）提出，是一很高效的规则。线性归结，由Loveland（1970）和Luckham（1970）独立提出，后由Anderson和Bledsoe（1970），Reiter(1971)，Loveland(1972)，Kowalski和Kuehner(1971)加以改进。语义归结、锁归结、线性归结都是完全的，它们中每一个都可单独用来从一不可满足子句集导出空子句。

2. 十年来的主要证明器

*)863计划软件生产自动化课题的部分资助项目

2.1 美国Argonne国家实验室的定理证明器

Argonne是世界上ATP(自动定理证明)研究的主要中心，由L·Wos教授主持工作。Wos曾获首届ATP"当代研究"奖，现任国际自动推理学会主席，《J. of Auto. Reas.》杂志主编。

在这个研究中心，诞生了AURA(Wos, 1981)，ITP(Lusk, 1984)。1986年，Butler实现了一个比ITP快100倍的新系统，它采用了Prolog中的实现技术和多处理器、可结合-可交换一致化、及数据库索引技术。

McCune与Boyer合作，实现了交互式证明检查器，用以证明Gödel公理化集合论中的一些基本数学定理。

Argonne小组将ITP广泛用于ATP研究，如证明定理、解决数学中开问题、验证软件和硬件等。ITP采用支架集策略、等式归结、归约和类含检查等技术，使用了完善的数据结构，其索引技术允许对项快速存取——用来与给定项进行一致化，使用了一个复杂的估值函数来决定归结步的优先次序，并提供友好的界面支持。ITP用Pascal书写，便于移植。现在ITP已分发到世界200多个地方。

2.2 KLAUS自动演绎系统

本系统由Mark Stickel (美国SRI)主持研制，它基于连接图方法，实现了如下ATP技术：

- 推理方向的控制。
- 允许一单独归结步包括整个"理论"，如重写(归约)，可结合-可交换一致化，多类一致化，分类分层等，提高了效率。
- 为了实现重写规则集，提出了Knuth-Bendix算法。
- 用估值函数作为优先控制机制。
- 包含一个PROLOG技术定理证明器，使用模型删除型类PROLOG线性搜索，搜索采用迭代加深策略。

Stickel证明了一大批标准ATP测试定理及数学中的定理。

2.3 Margraf Karl反驳证明器(MKRP)

此系统是德国Kaiserslautern大学N. Eisenger, H. J. Ohlbach, J. Siekmann等人多年来开发出来的一个能力很强的系统，它使用连接图方法，子句集中每一可能的推理步均表示为图中的连接，在执行了一组推理步之后，删去不相关和冗余的连接。这是算法效率的来源，同时也是问题的来源：对带推理限制策略的连接图归结，完全性定理不成立。

MKRP中，可做多类一致化，其中的"终止测试模块"能快速测试一集子句的反驳能否直接完成。

Kaiserslautern小组目前正在开发一个后继系统HADES(高级自动演译系统)。除了别的特性，它将在连接图中包含做为原子推理步的高层连接，用来证明半群和自动机教材中的所有定理。

2.4 Munich定理证明器

这是由德国W. Bibel主持的Munich小组在ESPRIT计划中实现的类似PROLOG的定理证明器。这个证明器基于Bibel连接方法，输入必须预处理成子句形式，使用有界深度搜索，系统对一阶逻辑完全。为了提高效率，采用了子句集特别归约。目前，正在研究的特殊硬件包括：为存取连接的可结合存储器和高度并行多处理器。

2.5 美国North Carolina大学的SPRF

由Plaisted主持研究的SPRF (Simplified Problem Reduction Format)，将PROLOG扩充至整个一阶逻辑。它采用Gentzen型公理系统，其中相继式形如"$\Gamma \rightarrow L$"，这里Γ是正文字集，L是一个正文字。

SPRF的代码非常清晰，大约15页PROLOG程序，搜索策略采用迭代加深，深度受限，子目标的解被存储。为了完全性，对非Horn子句采用特别的分裂规则。SPRF的速度与ITP及KLAUS自然演译系统等主

要证明器相当。

2.6 英国Edinburgh大学的LCF

由Milner, Gordon主持研究的LCF (Logic for Computable Functions) 是一个在类型λ演算中定义的验证计算函数性质的大系统，它在ML(Martin-Lof) 中有效地实现。

LCF已用来验证数千条标准数学定理，现在已由Larry Paulson扩展到高级演译上。

2.7 美国Texas大学的BM证明器

BM(Boyer-Moore)证明器是一个验证递归函数性质的大系统。它的核心是数学归纳法，但数学归纳证明过程须与简化证明过程配合使用。

BM定理证明系统的简化过程本身就是一个小的定理证明程序，简化程序包括重写规则简化、删除、使用等式、推广和删除无关项等技术。每种技术本身构成一个小的证明子程序，简化过程依次调用这些证明子程序。

如果上述简化过程都没能将要证的公式简化为T，则BM定理证明系统将自动转向对公式的归纳法证明。纳归证明使用的搜索策略是BM定理证明系统的核心，其关键问题是怎样得出对所证公式适用的归纳方案。

2.8 吴-周几何定理证明器

1978年，中国科学院系统所吴文俊教授在“中国科学”上发表了“初等几何中定理证明的判定问题和机械化”的文章，在理论上做了开创性工作。周咸青原是中科院硕士，后留学美国Texas大学，做了平面几何证明器，能够证明约2000条定理。

吴-周证明器的过程是，将几何定理的前提和结论转化为等式，证明结论可由前提导出。当然，这个方法仅适用于前提和结论均可表为等式，而不适用于含不等式情况。此外，这方法亦可拓广到微分几何等领域。

吴文俊的思想经周咸青在计算机上实现之后，在世界上产生了极大影响。相应地，1990年，在科学院系统所成立了数学机械化研究中心，吴教授任中心主任。1991年3月—6月，在南开大学数学所举办计算机数学学术年活动，旨在以数学机械化为重点，力求让我国更多的人了解和掌握此方法和思想。

2.9 美国Texas大学的MCC证明器系列

以Bledsoe为首的小组先后研究了如下证明器：自然演译型证明器IMPLY，分析中的一般不等式证明器，基于语义的分层证明器SHP，类比证明器，数论中的证明检查等。

2.10 Tableaux：模态逻辑的一般定理证明系统

此系统由法国Catch主持研究，用于命题模态逻辑，它对各类模态算子和一大类模态逻辑（包括时态，认识，动态逻辑）提供了一个统一环境。

该证明器的判定过程采用语义表列方法，这里的语义表列与传统的语义表列略有差别。本系统用VM/Prolog语言书写，在IBM-370机上运行，目前，对大多数例子来说，效率合理。这个系统为学习逻辑提供了工具，相对于采用语法演译的系统而言，本系统采用语义方法，独树一帜。

3.关于证明策略

研究自动定理证明，必须考虑搜索策略、化简、语义、抽象和类比、发现相关公理以及一些一般软件工程命题。通常一个证明方法提出之后，首先考虑其完全性。人类数学家在做推理时，并不是完全的，但这不成为问题。进一步，一些不完全的特殊方法也有一席之地。一般，我们不刻意追求一定理证明器的完全性。但是，不完全证明器在做简单问题时失败，就令人为难了。

一阶谓词逻辑的定理证明方法，很大一部分是基于Herbrand定理的反驳法。反驳法早期最有代表性的工作是Prawitz分析方法和Robinson的归结推导方法，它们分别要求将给定公式转化为析取范式和合取范

式，其缺点是，大量使用分配律，公式体积膨胀，证明过程不自然。

近十年来，发现了非范式的分析法（一般配对方法），非范式的推导方法(非子句归结、嵌套归结、特殊关系推导规则）以及分析方法与推导方法相互融合的方法（语义图上的路径归结）。

自动定理证明领域，尽管取得了许多成就，但仍有许多问题尚待解决。1988年，Wos著书“自动推理：33个基本研究问题”。书中列举了关于策略、推理规则、归约、类含检查、知识表示、综合方法、逻辑程序设计、自分析、情况分析法、归结法、定义的推广或简化、发现新定理、方法的选择以及类比推理等33个问题。我们将在附录中予以列出。

4.结束语

自动推理（包括自动定理证明）领域，仍处于初级阶段，需不断试验。我们针对Manna，Pnueli提出的线性时态逻辑，在386微机上用Turbo-PROLOG语言实现了“命题时态逻辑定理证明系统”[1]。该系统具有良好的界面，输入公式与手写公式一致，若输入不符语法，则光标自动定位在出错位置。定理证明全过程分步显示在屏幕上。

附录 33个基本研究问题

·策略

1）如何重新定义支架集策略或者扩充之，使其对层为1的子句均有能力处理？

注释 支架集策略是一种限制推理规则的应用策略。输入子句标志为0层子句，0层子句的生成子句称为1层子句。

2）采用什么样的策略来控制模参(在项一级)，达到与支架集策略控制面向文字的推理（在子句一级）相同的程度？

注释 推理规则模参推广了通常的等式替换概念。特别，这个规则允许一推理程序，通过发现适当的等式替换，以一子句改变个体项。另一方面，同UR归结中的规则一样，用来删除文字。模参是面向项而不是面向文字的。

3）是否存在一限制策略，在能力上比支架集策略强？

注释 禁止一自动推理程序检查搜索信息空间的所有部分之策略称为限制策略。支架集策略可视为一限制策略。

4）关于诸如群论中“交换子定理”问题，是否存在一策略，允许一推理程序用超归结，且跟模参的效率相同？

注释 自动推理中基准检测问题之一是群论中交换子定理：$\forall x(x^3=e)\rightarrow\forall x\forall y([x,y],y]=e$，这里$[x,y]$表示$xyx^{-1}y^{-1}$。以模参为推理规则，在IBM195机上，花2秒多CPU时间，可得该定理的证明。用同样的机器和程序，但以超归结推理规则代替模参，其证明需要约100秒CPU时间。

5）除目前的语法规则外，还有什么准则可用来在选择子句中指导推理程序——应用特殊推理规则，而不是简单地发现一子句，其中包含一带适当标记谓词的文字？

注释 此问题的解决会明显提高推理程序的效率。语义准则，一旦被应用，必然比语法准则强得多，例如，支架集策略的能力，之所以成为可取，归因于对语义准则的依赖。

·推理规则

6）是否存在一推理规则，与模参同样执行，但它是从不等式而不是等式出发做推理？

7）是否存在推理规则，对集合论的执行如同模参对等式的执行一样有效？

注释 集合论在许多数学领域以及别的研究领域起关键作用。一推理规则允许自动推理程序不必应用相应公理，而直接存取集论变量，极大地改进了程序的效率。如果这样的推理规则能有效地实现，则数学中许多深奥的和至今未解决的问题可借助自动推理程序来解决。

8）如何给超模参(hyperparamodulation)一个适当的定义以避免产生所有模参子（paramodulants)？

9）使用超模参推理规则时，什么准则导致推理程序附加新的核(nuclei)？

10）在解释关于诸如群论中“交换子定理”的超归结和模参不同行为时，什么性质必须给出，什么性质可以缺省？

11）当效率做为选择标准时，一个问题及其表达的什么性质支配推理规则间的选择？

·解调

12）在使用超归结、UR归结、超模参的过程中，允许使用解调(demodulation)的则准是什么？

13) 除了深度准则以及产生一完全归约集准则外，用什么准则来决定哪个等式单位子句附加到解调子表中？

14) 什么样的理论适合于超出变元和文字界限的解调——这样的理论与解调类似，或者与完全归约集类似——同一些谓词代替另一些谓词，用一些文字集代替另一些文字集？

15) 是否存在准则，能精确预测完全归约集的大小？

16) 一个理论接纳一完全归约集，必须满足什么样的准则？

17) 一理论满足什么准则，保证一完全归约集不存在？

·包含

18) 对输入子句和生成子句，应用什么样的准则来决定使用何种类型的包含(subsumption)——比如，特殊单位子句、基、简单等式等？

·知识表示

19) 用什么准则来有效地选择谓词和函词概念，以此表达各种各样的信息？

20) 什么准则可用来有效地选择使用或避免使用等词？

21) 什么准则可用来删除函词，代之以适当的谓词或常元，这些常元是用来重新命名原先由函词"命名"的实体？

·综合方法

22) 给定一个待解决的特殊问题，——记住表示、推理规则和策略间的紧密联系——什么样的元规则使得同时选取最好的表示，将要用的最有效的推理规则以及最有力的策略来控制这些推理规则？

23) 在子句表示和自然演译表示（以及相应的推理规则和策略）之间是否存在对应，使得基于两种方法的推理程序以本质上相同的方式解决一给定的问题？

24) 用来阻止一推理程序产生一已存在子句的首要准则是什么？

·逻辑程序设计

25) 判定一问题是否适合于逻辑程序设计，而不需要使用通用自动推理程序的首要准则是什么？这个准则用来将问题分类成专用算法可解的问题和需要新信息进行不确定搜索而求解的问题。

·自分析

26) 在解决某一给定问题中，什么准则可用来使一推理程序做自分析，确定支架集必须做调整使解题效率迅速提高？

27) 在解决某一给定问题中，什么准则可用来使一推理程序做自分析，确定权必须做调整使解题效率迅速提高？

·其它问题

28) 考虑某给定问题时，使用什么准则来决定采用情况分析法，对这样的方法，使用什么准则来选择所考虑的情况？

29) 使用什么样的准则来选择性质，对此采用归纳法来讨论？

30) 使用什么样的准则来导致一推理程序推广或简化定义？

31) 确定什么性质，自动推理程序借以有可能发现有趣的新定理(而不是证明已有定理)？

32) 当一特殊问题正在被求解时，使用什么准则来选择已解决的问题，以此确定当前问题要使用的方法？

33) 考虑到人们在求解问题中使用多种推理，对一问题来讲，应用什么样的准则来使一程序做类比推理？

主要参考文献

[1] 贲可荣，陈火旺，命题时态逻辑定理证明新方法，软件学报(已录用)

[2] Bledsoe, W. et al., A survey of automated deduction, Exploring Artificial Intelligence, Survery Talks from the Notional Conferences on AI, Howard E.S. ed. 1988

[3] Plaisted, D., Mechanical Theorem Proving, Formal Techniques in Artificial Intelligence, 1990

[4] Post, S., & Sage, A.P., An Overview of Automated Reasoning, IEEE Trans. Syst. Man Cybern., Vol. 20, No.1 1990

[5] Wang Hao, Computation, Logic, Philosophy, A Collection of Essays, 1990

[6] Wos, L., Automated Reasoning, 33 BASIC Research Problems, 1988

[7] Wu Wen-tsun, Automatic of Theorem-Proving, Mathematics-Machanization, Research Preprints, 1990

及1991年于南开大学数学所举办计算机数学学术年活动的相关资料。