

能力提升

高质量读研以及撰写学位论文的若干建议

邓 鑫 刘宇航

中国科学院计算技术研究所
中国科学院大学

关键词：研究生培养 学位论文写作 质量 效率

引言

我们生活在一个节奏快、注意力稀缺、时间短缺的时代，研究生在追求获得学位、寻找工作机会，研究生导师在写申请书申报项目，诸如此类。而培养人才、撰写学位论文都是需要慢功夫的长远事业，既急不得，也马虎不得。本文第一作者邓鑫攻读硕士学位期间的工作已经应用到工业界实际系统中，通过了较为严格的测试，还发表了两篇 CCF 推荐 A 类国际学术会议 / 期刊论文^[1,2]，这是比较不容易的，原因有三：第一，计算机体系结构领域的研究大部分属于“重工业模式”，应用到工业界并非轻而易举，发表 A 类论文也并非易事；第二，邓鑫是专业学位硕士生，本科毕业于重庆邮电大学，起点并不太高，基础并非足够扎实；第三，邓鑫是其导师刘宇航独立指导培养的首位硕士生，课题组力量并不特别雄厚，导师的指导经验并非足够丰富。

相对那些资深的研究者和积累深厚的课题组，本文作者可能班门弄斧。但我们希望不揣浅陋，正值毕业季，把读研的一些感悟以及撰写高质量学位论文的若干实操性建议分享出来，供大家参考，若得少许裨益，也倍感欣慰。本文第 2 部分由邓鑫撰写，第 1 部分和第 3 部分由刘宇航撰写（卜东波研究员

给予了大量帮助）。

读研的感悟

在中国科学院计算技术研究所读研的三年，是我人生中一段宝贵的经历，使我在技术和心性上都得到了锻炼和提高，在知识、技能上收获了很多，也走了不少弯路，踩了不少“坑”（pitfall），在此分享一些让我印象深刻的经验和教训。

明确目标、分解目标

首先需要明确研究的最终目标，并将其拆分成里程碑式的小目标。不清晰的研究目标可能会让我们多走很多弯路。比如对方法论的研究，需要对系统进行一定的抽象，而且要避免过多地考虑一些不常见的特殊情况而导致研究陷入一些工程细节。对于这类工作，可以先针对典型样例完成验证性的工作，证明方法的有效性，再逐步完善。在确定研究目标后，研究过程中的技术路线也是多样的，需要仔细思考后将其拆分成里程碑式的小目标逐个击破。

拆分成里程碑式的小目标并不简单。拆分的过程需要我们对研究目标进行深入思考，也需要预先设想不同的技术路线可能要面临的各种挑战，以确保拆分

的每个小目标能够有效帮助我们完成最终目标。

思路清晰、抓住关键

研究者需要避免在细枝末节的问题上花费太长时间。在研究的过程中会遇到各式各样的问题，遇到问题时不要急于解决，而是要反思自己的研究目标，想一想是不是真的有必要去解决。

以图1为例，我们在选择技术路线时可能选择了看上去更加简单的经过里程碑A和里程碑B的路线来实施。但是在实施的过程中遇到了问题1，要解决问题1又要解决问题2……在这种情况下，就要避免陷入解决无关紧要的问题，可以尝试跳出既定的研究路线，重新思考是否还有其他灵活的方法可以绕里程碑A。

在之前的一次经历中，为了测量更多延迟敏感型应用的尾延迟数据，我们查阅资料发现，ISCA'16的一篇论文中提出的 Treadmill 工具开源了一个相对于其他工作更加精确的延迟测量工具。我们自然地将其中的一个小里程碑定为使用 Treadmill 工具测量应用的延迟数据。在实践的过程中，由于该工具依赖了接近 10 个我们并不熟悉的库，并且随着版本的更新，工具与库的接口已经不再兼容，并且也无法查到当前的工具版本能正确依赖过去哪个版本的库，导致在安装过程中刚解决了一个问题又

出现了另一个问题，最终花费了两个星期仍然没有成功安装。

现在回忆起来，当时应该跳出当前遇到的编译问题，思考是否一定需要使用这个工具才能进行后续的研究，切换成其他精度稍差的测试工具会不会影响实验的结果，或者不测量这个应用，更换其他应用是不是有更加快捷的方法。因此，在研究过程中，应避免陷入解决琐碎的问题，要注重效率，关注工作的核心创新点并优先解决其中的主要矛盾。

定期整理、规范记录

做研究时要定期整理并规范记录研究的数据和代码。在研究过程中，往往会有很多中间结果，定期整理这些中间结果是有意义的。

一方面，这些数据在论文或结项报告中可能会用到。根据我自身的经历，用到这些中间数据的时候往往是在几个月后或一年后（比如论文投稿需要补充数据时）。如果没有做好记录，需要使用时再重新梳理代码生成数据反而需要花费更多的精力。这些提前整理的文档或数据以后可能会意想不到地用在毕业论文中，减轻以后的工作量。以我自身为例，在导师督促下写的一些文档，很多都直接应用在了毕业论文中。

另一方面，整理数据和代码的过程可以帮助我们梳理研究的脉络，把研究的整个“故事”讲得更清晰，在梳理的过程中能够及时纠正当前研究过程中存在的问题。除此之外，这些定期的整理，也能够帮助我们缓解服务器中长期不整理的数据导致磁盘空间不足的问题。

比如在我们在表述和证明云计算中的 DIP 定理¹的工作^[2]中，起初是导师先整理了一个论文版本，我在写毕业论文时对该部分

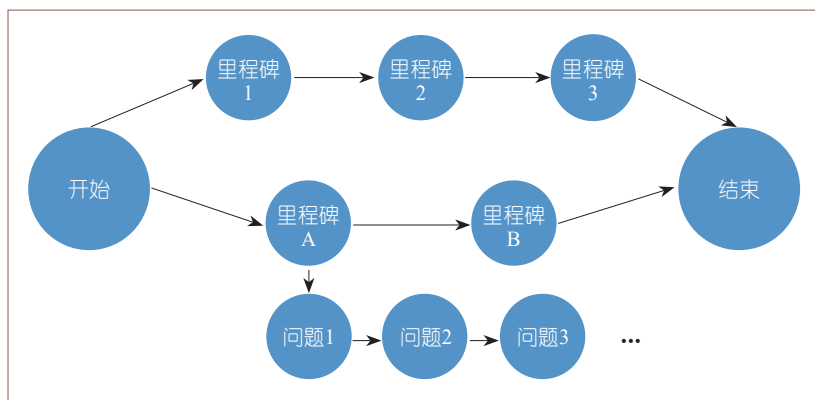


图1 技术路线的选择

¹ 在云计算领域，DIP 定理是指：区分（Distinguishing）、隔离（Isolation）和优先调度（Prioritizing）是数据中心应该同时具备的资源管理能力，以保证数据中心上各种类型的用户从整体上获得尽可能理想的用户体验。

补充了实验数据后再次进行了整理,使该部分内容更加完整,故事的来龙去脉更加完善,但是当时整理的内容仍然存在一些问题,比如文字还不够精练,读者较难理解。最后在写期刊文章时,我再次对该内容进行了整理,将该部分的内容写得更加准确、精炼且易于理解。如果这部分内容不经历如此多次的整理,可能无法精练准确地表达出想表达的意思。同时,在整理和思考的过程中往往也会有新的思路和收获。

撰写学位论文的建议

几千年来,文章一直是文化的重要载体;近代以来,论文是科技传播的重要媒介。以规范的形式高效率地传递信息,是对科学技术从业者的一项素质要求。

研究生在撰写学位论文时,常常犯一些错误,既有格式错误,也有内容错误。本节列举一些实操性的建议,以求减少师生检查论文之辛劳。

态度

需要强调的是:方法和手段不是第一位的,写好论文的基础是认真而诚恳的态度。作为母语非英语的人,笔者发现一个值得注意的现象:在英语中,科学(science)与良知(conscience)竟是如此相像和接近。“con”表示“共同来做”,共同致力于科学就是良知。科学需要唤醒良知。门捷列夫²说,没有测量,就没有科学。笔者想表达的观点是:没有良知,就没有科学³。笔者的一位同事在研究计算机领域的基准测试学或评价学,笔者认为这是一项重要的但不容易的工作。笔者的看法是:(1)任何决策(包括智能体或人类的所有活动涉及的决策)都需要评价(包括来自自身的或外部的评价);(2)评价本身需要评价,评价的评价也需要评价,以此递归;

(3)评价的终极基准就是良知。

研究者从事研究活动的全过程中,包括撰写学位论文的过程中,良知都在发挥牵引和自我评估的作用。追求真善美、信达雅,是良知在发挥作用。态度是良知的外化,是能否避免或纠正错误的关键,实干是最好的方法学。如果没有端正的态度,就会出现无穷无尽的错误(例如广受诟病的“图片误用”)。如果没有实干,一切都将是纸上谈兵,而“纸上得来终觉浅”。

标题

摘要是全文的浓缩,标题是摘要的浓缩。标题要准确。定语位置要恰当。论文的英文题目中常出现“Research on xxx”,这是中文题目“关于xxx的研究”的硬译,殊为不妥。哥伦比亚大学建议的规范写法是“A Study of xxx”,或者直接陈述贡献,如“Algorithms for xxx”“On the Intractability of Genome Sequence Assembly”。

摘要中的主语该用什么?

陈述贡献时,主语不要用“我们”,应改成“本项研究”或者“本文”。与期刊/会议论文(paper)不同,学位论文(thesis)是个人作品,供学位评定之用,因此具有浓厚的个人色彩。为厘清其他个人或集体的贡献,学位论文中的前面都会包含一个声明,比如“学位论文系本人在导师指导下独立进行研究工作取得的成果;对本论文涉及的研究工作作出贡献的其他个人或集体,均已在文中以明确方式标明或致谢。”因此,在学位论文中陈述贡献时,不能用“我们”。

解决方法:以“本文”或“本项研究”代替“我们”;在英文中,使用“Our research shows...”或“The conclusions made from this analysis are...”替代“we”。需要注意的是,这不是说通篇不能使用“我们”。“我们”有两层含义:一是表示“作者及其合作者”,即

² 门捷列夫,俄国化学家,发现了化学元素周期律,曾担任俄国度量衡局局长。

³ 笔者先有自己的这一观点,后来偶然了解到门捷列夫的观点。一个可能值得进一步思考的问题是,这两个观点之间,或者说,测量与良知之间,是否存在某种深刻的联系。目前笔者的看法是,测量需要度量衡,而良知可能是终极意义上的度量衡。

“With the assistance of my colleagues, I designed ...”;二是表示“作者和读者”，即“the readers”。在采用第二个义项时，可以使用“我们”。

摘要中如何描述贡献？

摘要反映了一个作者的总结凝练能力。描述文章贡献时，建议采用“冒号式、分点陈述”格式（这也是 ChatGPT 的常用格式），使用“提出了一种 / 开发了一套 / 构建了一个”等说法。例如：

本项研究取得了如下成果：(1) 提出了一种 $\times\times\times$ 算法（详细陈述目标、基本思想、性能等）；(2) 设计了一个 $\times\times\times$ 系统（详细陈述目标、基本思想、性能等）。

绪论

绪论是文章之始，开篇最为重要，要讲清楚为什么做这个研究、解决什么问题、现状是什么、基本思路是什么、结果是什么、全文的布局等。绪论中要写明研究动机（motivation），并减少与摘要重复的内容。在研究的不同阶段，动机的阐述所占的比例不同。

如图2所示，平时做研究时，花费最多的时间和精力在“解决问题”部分上。我们需要尝试各种各样的思路，编写和调试大量的代码，处理杂乱的数据。如果效果不好，又要从头再来一遍。这个重复的过程很艰辛。因此，在最终展示自己的工作时，人们也会不由自主地展示大量艰辛的工作过程。但这在论文撰写或口头汇报时是不可取的。

如何解决问题固然重要，但如何发现问题、解决这个问题有什么意义、最终得到什么样的结论，这些问题同样重要。因此，在论文正式部分开始之前，需要花大量篇幅写“绪论”和“相关工作”（有

些学科也把这个部分称为“文献综述”章节），在最后的章节要分析数据、总结结论和进行展望。在论文写作中，发现问题、解决问题、得出结论这几部分都很重要。

在进行口头表达（oral presentation）时，情况又发生了变化。口头表达的时间有限，知识基础各异的听众在短时间内理解大量技术细节也几无可能，应该着重介绍“发现问题”和“总结结论”部分。在“解决问题”部分，需要提炼出解决方案中最精华的部分，最好不要陷入到琐碎的技术细节中，否则读者或听众可能会很快失去兴趣。

句子/段落避免过长

单个中文段落尽量不要超过10行；一般而言，超过10行的段落会超出段首句的“管辖”，造成阅读困难。一句话不能过长，有些人写文章，一段只有一句话，“一‘逗’到底”，一方面反映了语文修养不够，另一方面反映了逻辑思维训练不够。标点符号的使用，影响的不仅仅是文采和“雅”，还影响思维逻辑，影响“信”和“达”。

解决方法：一旦句子过长，就立即分段或者采用“总—分—总”式结构，以结构引导读者阅读。将一句话分成若干点逐点表述，每个点都通过一句话描述，一句话说一个意思。

关于使句子逻辑性更强、更便于理解的方法，这里介绍几个小技巧，写作的入门阶段可以使用这些小技巧刻意进行练习。

有意识地多使用连接性的词语，如“因为……，所以……”“为了……，需要……”“虽然……，但是……”“首先，……；其次，……；再次，……；最后，……”“第一，……；第二，……”。当把这些词语写出来的时候，大脑便会自动地开始思考如何组织词语才能符合这些连接性词语规定的逻辑，进而让写出的文本更有逻辑性。

能用句号就用句号。有意识地多使用句号，可以让句子更短，从而间接地促使我们用尽量简洁的语句表达想表达的意思。

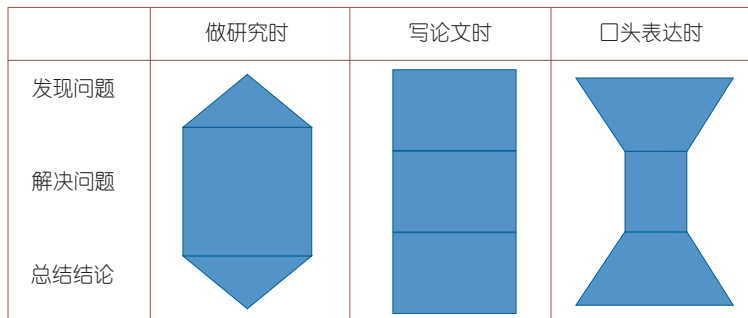


图2 研究的不同阶段需要不同的时间分配

在写每一句话前，先在脑中叙述一遍自己想要写的句子。对该句子进行整理后，再形成书面文字。边想边写的后果通常是文本质量不高、逻辑不清晰、“口水话”的现象特别严重。

英文缩写

对于首次出现的英文名词，应采用“中文名（英文全称，缩写）”的格式；另，拉丁文要用斜体。比如大肠杆菌，应该写作“*E. coli*”，中间有空格。中国科学院大学给出的相关规定摘录如下：特殊名词或新名词应在适当位置加以说明或注解。双名法的生物学名部分均为拉丁文，并为斜体字。采用英语缩写词时，除本行业广泛应用的通用缩写词外，文中第一次出现的缩写词应该用括号注明英文原词。新的外来名词应用括号注明英语全称和缩写语。

力戒中英文混杂

例如“benchmark”，需要翻译成中文，不可中英混杂。另，常见的“Transformer”等都有规范译法。

避免含混、指代不清

不要出现易混淆词汇以及意思不明确的句子。例如“序列”，要讲清楚是“蛋白质序列”，不要以为读者能够靠上下文自然补足。再如卜东波研究员课题组某学生毕业论文中某节标题为“3.2.1 XY-Bind”，卜老师指出：这里直接说XYBind，读者会看不明白；应该在前面加上“蛋白质－肽段相互作用预测算法XYBind的整体流程”。

公式

公式不能使用Python代码代替。公式要有编号，正文要有对公式的说明和使用。

表格的表头

表格的表头要用中文，不能用英文；数字一般情况下要精确到小数点后3位，采用右对齐方式，以利于比较。中国科学院大学相关规定摘录如下：表内同一栏数字必须上下对齐。

表格的格式：表格不要有竖线

中国科学院大学相关规定摘录如下：表格尽量用“三线表”，避免出现竖线，避免使用过大的表格。

图表的标题

图表的标题要避免含混不清，不能指望读者根据上下文猜测、补全标题的含义，而应该是“含义自明的”，要突出与文章的关联；另外，图表的标题结尾不能出现标点符号。例如：某个图的标题直接用“Social-LSTM”则比较突兀；前面应该加上短语以明示，比如改为“个体相互作用检查算法 Social-LSTM 中的社会力池化操作”，或者改成“基于……（社会力）模型……（任务）预测方法”。

中国科学院大学的相关规定摘录如下：各章标题中尽量不采用英文缩写词，对必须采用者，应使用本行业的通用缩写词，标题中尽量不使用标点符号。图应具有“自明性”，即只看图、图题和图注，不阅读正文，就可理解图意。每一幅图都应有简短确切的图题，连同图序置于图下居中。此外，缩略语常常总结成表，放到论文正文之前或之后，以利于读者查找；在缩略语表中，缩略语按照在论文中出现的顺序排列。

学位论文是“thesis”，不是“paper”

不能用“This paper presents”，而用“This study/thesis presents”。时态是现在时，语态是主动语态。

段首句要统领全段

每一段开头是总括句，要能够有效地统领全段。

力戒指代不清

不要用“其”“它”指示某种算法或者工作；最好直接用名字，以避免指代不清；不要用“效果”“性能”这种含糊不清的“大词”（big words），直接用“准确度”等语义指向清晰的词。

一行不能以标点符号、数字开头

不能用标点符号开头，不能用数字开头；章节不能用图或者公式开头。

标点符号

中文中的顿号不可用逗号代替。中文句子内部的并列词，应该用全角顿号分隔，而不用逗号，即使并列词是英语也是如此。用LaTeX排版时，左引号用“”或“”。

避免滥用被动语态

滥用被动语态是一种常犯的错误，比如“2022年，一种新的蛋白质结构预测算法被提出”；正确的写法是：“2022年，Jumper, et al. 提出了一种新的蛋白质结构预测算法”。

数字

每个有单位的数字都要给出单位，没有单位的数字要给出说明。

算法描述

算法描述由输入、输出、参数说明以及算法主体组成，缺一不可。算法描述要写明输入、输出，使用到的常量要解释并给出取值，代码片段也要给出其中的变量说明。形式化描述完算法后同样也要用自然语言描述清楚。

这里强调一下“算法可以没有输入，但是一定要有输出”。有些时候，我们会使用面向对象的语言实现一个算法，其中包括几个数据结构以及应用在这些数据结构上的几种操作（如初始化、更新、检查等）。某些操作对应的函数可能没有输入，因为它操作的是类中的某个数据成员。在这种情况下，当书写算法的伪代码描述时，需要把操作的那个数据结构作为算法的输入，把操作之后的结果作为输出。切忌直接对照代码翻译算法描述而忘记了算法的输入输出。

数学公式

用数学符号或字母表示各个量，此时不要出现中文。

引用

文字内容对表格的引用尽量不要跨很远。**解决方法**：绘制新的图表，或以文字形式描述所需图表内容。

结构图

论文中可以插入几张精心绘制的图片，加强论文的可读性。如绘制文章结构图、原型系统的框架图、主算法的流程图等。

专有名词

论文中存在的主要问题有专有名词用词错误、论文前后用词不一致等。**解决方法**：将论文中引用

的专业术语以及提出的概念在论文中用列表给出。通读论文，检查是否存在同义不同词的情况。专有名词有中文翻译的给出中文翻译。

句子成分齐全，表意清晰

每一个句子语法都要完整，不要“缺胳膊少腿”。学位论文没有篇幅限制，无须刻意省略。**解决方法**：减少不必要的省略，使每一句话单独拿出来都可以被读懂。

举一个实例，修改之前的句子是“预测正确计数器增加1，预测错误减少1。”这句话没有表达完整。修改之后的句子是“在猜测正确的情况下，置信计数器增加1；否则，置信计数器减少1。”

图表清晰

在LaTeX插入较宽的图表容易导致图表的文字模糊，难以辨认。**解决方法**：表格竖排或调大图表字号；或者插入PDF格式的图表。另外，最好将全文的图片都设定为相同的宽度，否则插入到LaTeX后，可能会由于图片放大或缩小导致图中的文字字号不统一（即使在绘图时使用了相同的字号）。

章节开头/结尾

每一章的开头都要有导语，总领全章；每一章的结尾都要有结语，总结本章的工作；章节按“总—分—总”的形式撰写。

章节标题

章节标题不宜过长，要精简凝练。举一个实例，修改之前的标题是“第6章 基于访存模式向量的自动构建基准测试程序集的方法以及对多个领域的大量应用的访存模式分析与比较”。作为一章的标题，显然过长。修改之后的标题是“访存模式分析器的应用示例”。

实验结果

要详细给出对比的项目来自哪里，描述结果时遵循问题导向，思考数据验证/解答了什么问题。避免不分析数据，或者只进行表面化的“分析”。

语法词法

区别“的、地、得”，避免出现错别字，做好校对（proof reading）工作。“的、地、得”的用法属于小学语文的内容，但是相当一部分研究生不会

正确使用。“的”通常用作形容词后的定语标记，用来修饰名词，表示所属关系或特征，比如“访存密集型的应用”。“地”通常作为状语的标记，用来修饰动词、形容词或整个句子，表示方式、程度或场所，比如“剧烈地抖动”。“得”通常用作补语的标记，用来修饰动词、形容词或整个句子，表示程度、结果或条件，比如“增长得快”。

附录

论文和代码都要规整，可以在论文中加入附录，描述代码的工程实现结构、文件、代码量等信息，方便读者或后续研究者理解和延续自己的工作。

研究心得

论文末尾最好写一点关于“成果是怎样做出来的”的心得体会。确切地说，论文如果没有这部分，不是一个错误，这里只是一个倡议。研究生们经过三五年或六七年的思考和探索，终于写成了学位论文，以记录和展示自己的探索成果。然而，有过科研经历的人都清楚，展示“成果是什么”与展示“成果是怎样做出来的”，是两件截然不同的事情：（1）展示“成果是什么”，采用的模式基本上都是演绎推理，即“目的→方法→结果→结论”的逻辑链条，这是学位论文自有其严谨性和逻辑性之要求使然。（2）想要展示“成果是怎样做出来的”，相对来说有些复杂：做研究的过程，是尝试，是反复失败后的反思与调整，是陶哲轩所谓的“路径搜索”；是归纳逻辑，是跳跃性的联想，是波利亚（G. Polya）⁴所谓的“合情推理”；是猜测和验证，是迭代式改进，是迭代式改进中偶发的“相变”。

研究经历通常不是一个平稳的线性上升过程，而是一个非线性的阶跃过程：一开始，不免有些迷茫与徘徊；在某一时刻，因为某个原因，做了某种尝试，而拨云见日，势如破竹。因此，如果在论文中仅仅展示“成果是什么”，实则近于买椟还珠：后来人能欣赏到理论的高妙，而理论背后隐伏的摸索与徘徊却被忽略了；后来人能欣赏到性能的高超，

而性能背后隐伏的迷茫与灵机一动也被忽略了。有些学生讲述了自己的思考过程，单独作为一篇后记，或者放于致谢之中，感谢在这个过程中得到的帮助。故此，倡议诸位毕业生不妨写一写自己的科研历程，这既是对自己数年探索的总结，也是对后来者极有启发性与借鉴意义的“他山之石”，是对后来者的“授人以渔”，是一件有着长久价值的事情。 ■

致谢：感谢卜东波研究员及其课题组提供的宝贵素材，邓鑫的师弟陈泓佚、周嘉鹏等同学对全文进行了补充和校对。



邓 鑫

中国科学院计算技术研究所2019级硕士，腾讯微信事业部研发工程师。



刘宇航

CCF 高级会员，CCCF 专栏编委。中国科学院计算技术研究所副研究员。主要研究方向为计算机体系结构、高性能计算、存储系统、智能并发系统。liuyuhang@ict.ac.cn

参考文献

- [1] Liu Y, Deng X, Zhou J, et al. Ah-Q: Quantifying and handling the interference within a datacenter from a system perspective[C]// *Proceedings of the 29th IEEE International Symposium on High-Performance Computer Architecture (HPCA 2023)*. 2023: 471-484.
- [2] Liu Y, Deng X, Zhou J, et al. Suppressing the interference within a datacenter: Theorems, metric and strategy[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2024, 35(5): 732-750.

（本文责任编辑：刘宇航 郭得科）

⁴ 波利亚，美籍匈牙利数学家，历任布朗大学、斯坦福大学教授，法国科学院、美国国家科学院和匈牙利科学院院士，著有《怎样解题》《数学的发现》《数学与猜想》等。