

学术论文阅读与写作

Zaiwen Wen
Beijing International Center for Mathematical Research
Peking University
wenzw@pku.edu.cn

July 4, 2022

1 论文阅读

每周坚持写学习科研总结和文献阅读笔记:

- 总结上一周的工作进展
- 课题任务类型: 应用/建模, 算法, 理论分析, 数值试验等等
- 根据课题任务, 把任务分解成本周可以实现的小任务, 哪些是要学习的分析技巧/算法技巧, 哪些是要研究的内容。
- 理清课题主要贡献, 如何寻找课题切入点, 目前主要困难, 有哪些解决策略
- 坚持写作, 整理论文草稿, 读书笔记等等
- 及时总结讨论内容, 整合到科研总结里
- 主动跟导师或合作者更新和沟通进展。

1.1 总体原则

- **读厚**: 按照下面三步曲原则详细读论文, 不单要读懂论文本身, 遇到不懂的地方, 还需要去查找文献
 - 勤用google: 搜索词组或句子用" "扩起来, 指定文件类型filetype:pdf, 指定网址: site:。例如: "online optimization" filetype:pdf site: http://www.jmlr.org/
 - google scholar: scholar.google.com
 - 数据库: http://www.ams.org/mathscinet/
 - 文献管理工具: bibdesk (Mac OSX), JabRef。它们也提供了很方便的文献查询工具, 可以对接一些常用数据库直接查询
 - 坚持浏览arxiv等文献相关网站, 可以订阅邮件, 接受固定周期推送
 - * http://www.arxiv.org
 - * http://www.optimization-online.org
 - * Mathematical Programming
 - * SIAM Journal on Optimization
 - * SIAM Journal on Scientific Computing
 - * SIAM Journal on Imaging Sciences
 - * SIAM Journal on Matrix Analysis and Applications
 - * Journal of Machine Learning Research
 - * NIPS 网站

- * ICML 网站
- * 其它一些相关的杂志网页
- * 相关领域研究人员的网页
- * 课程网页，如：<http://suvrit.de/mit/optml++/index.html>
- 读书报告：按读文献不同阶段准备读书报告，不断完善，需要回答三步曲的问题。不偷懒，不同部分可以用在自己学术论文，学位论文，综述等等
- 精读与略读。其中略读是精读步骤的简化
- **读薄**：按照下面三步曲原则详细读论文，要能收回来
 - 读完之后一定要简要及时总结。一两句话的概括，一段话的概括，这些在自己写文章时的introduction等地方用到。
 - 在一定积累之后，可以找出文献直接的区别与联系，寻找自己课题的指导
- 细心，有耐心，持之以恒
- 科研遇到问题怎么办？
 - 放松一下：体育运动，看电影，找个好吃的，找朋友侃大山...
 - 找找相关文献，参见“读厚”原则，运用泛读方法
 - 列提纲，分解问题，哪些能解决，哪些不能解决，哪些需要解决，问题归类，从熟悉和简单问题入手
 - 问题化简，简化简化再简化成熟悉的问题解决之，然后一点点添加复杂度解决之，如此反复
 - 参考“阅读三步曲”里的“总结，思考与扩展”
 - 尝试该课题的其它事情，比如理论上困难先做数值实验，数值实验有困难先做理论问题。
 - 尝试考虑其它问题

1.2 阅读三步曲

- **初读/粗读**
 - 粗略浏览，顺序不一定是从头到尾，先读熟悉的，不熟悉的略读或跳过
 - 关注主要结论，问题的描述，主要定理，主要算法，主要数值实验
 - introduction 往往比较抽象，可以略读，后期再推敲。写作的时候introduction 通常是最后写的，读的时候也可以反过来
- **仔细阅读和思考**
 - introduction：研究什么问题，文献里已有哪些方法，这些方法存在哪些问题，作者的理论或方法是哪些，作者的主要贡献是哪些
 - abstract, introduction, conclusion 里claim了什么主要结论？这三部分都是简略的概括文章研究的问题以及贡献，但详细程度不太一样，侧重点分别有哪些？
 - 正文部分：
 - * 所研究问题的详细表述和定义，问题的基本性质（还包括文章里没有写的），比如（最优）解的性质，存在性，唯一性
 - * 理论文章：有哪些理论问题，主要有哪些假设，理论上的主要结论，证明的主要技术。任务：读明白，能写出证明
 - * 算法文章：算法motivation, 算法的具体描述，算法的理论性质，比如收敛性，收敛速度。算法的主要技术，理论证明的主要技术。任务：能推导出算法
 - 数值实验：程序如何实现的，算法的参数有哪些，测试算例是哪些，作者是否提供测试程序和数据？测试算例是如何设计的，主要的数值实验结论有哪些。任务：复制数值实验结果

● 总结，思考与扩展

- 研究问题: 问题描述和定义是否清晰, 简化形式有哪些, 扩展形式有哪些
- 理论文章: 作者有没有做到claim的理论结果, 假设是否合理, 假设简化情况如何? 假设加强情况下如何? 比文献里结果好在哪些地方? 哪些地方是否可以做得更好, 证明是否可以简化。是否可以推广到其它问题
- 算法文章: 作者有没有做到claim的理论结果, 算法的假设和适用范围, 比文献里算法好在哪些地方? 哪些地方是否可以做得更好, 如何改进算法, 是否有更好的算法, 是否可以推广到其它问题
- 数值实验: 数值实验是否解决核心理论问题和算法问题, 测试算例是否有借鉴意义, 是否可以解决其它问题
- 文章的精华在哪里? 我如何解决这个问题?

2 论文写作

● 语法问题

- 主谓宾原则。每一个句子都应该有完整的主谓宾语。负责的句子特别要仔细检查。如果复制的句子写不清楚, 先拆成简单的句子, 然后再考虑组装成复杂的句子。
- 初稿可以比较随意, 把能写的都写下来, 然后再一遍一遍的修改中慢慢改进。**主动的态度永远最重要, 不能等着老师找你修改**
- 学习文献的写法, 拷贝下来, 利用语法知识, 改写成自己所需要的形式
- 句子表达的多样化: 修改一些主要的词或短语, 修改句式, 修改从句的表达等等
- 多用google 查询, 参考读厚原则。不要怕麻烦, 开始阶段要敢于逐字逐句查询
- 多查数学字典
<http://bicmr.pku.edu.cn/~wenzw/MathEnglishDict.html>
- **严禁直接拷贝! 学会如何合理改写**

● 组织问题

- 总体原则: 有条理, 摆事实, 讲道理, 讲到位
 - 思考: 中文能写清楚么? 中文能表达完整么?
 - 写哪些内容? 参考阅读三步曲。下面是一个典型的写法(比较八股)
 - * abstract 是introduction 从不同侧面的缩写。
 - * introduction: 研究什么问题, 文献里已有哪些方法, 这些方法存在哪些问题, 我们的理论或方法是哪些, 我们的主要贡献是哪些, 本文的organization。上述每一个问题基本上分别一段
 - 正文部分, Preliminary: 问题的详细表述和定义, 问题的基本性质, 已有方法的一些综述等等
 - 正文部分, 理论文章: 理论问题表述, 引理, 定理的证明
 - 正文部分, 算法文章: 算法motivation, 算法的具体描述, 算法的理论性质, 比如收敛性, 收敛速度。
 - 数值实验: 算法的参数有哪些, 测试算例描述, 数值结果以图或表格形式表述, 主要结论。图中数字字体应该基本上与正文字体大小一致, 线条的粗细和marker 应该清晰可辨, 原则上是打印出来能非常直观
 - conclusion 也是introduction 从不同侧面的缩写, 与abstract略不同。conclusion 还可以有一些展望
 - reference 参考文献一定要规范, 学会使用bibtex, 管理自己的参考文献库
- 段落内部的组织。第一句一般为中心句子。句子之间推敲哪些先写, 哪些后写。哪些是原因, 哪些是结果。
 - 段落之间的组织。

- section 之间的组织。
- motivation 应该讲清楚。算法描述之前，定理和引理等理论结果描述之前应该有解释性语言，不能只是公式。
- 冰冻三尺，非一日之寒。从读书报告开始，一步步改进，每一次有改进
- **典型的一些语法错误**
 - a, an, the 的用法，什么时候加它们，什么时候不加
 - 标点符号。如果一句话是公式结尾，也应该有合适的标点符号
 - 待续....