

如何高效阅读机器学习顶会论文

1. 读书与读论文的区别

教材仔细读：经过长时间的打磨，是精品；书籍如CSAPP, ESL, MLAPP, GEB（最后这个花费作者一辈子心血）

论文要甄别：

- 顶会论文：cvpr, icml, nips, 等
- - 70%：在前人基础上做了一丁点改进，如改一下损失函数、调调正则化 -> **不值得花大量时间研读**
 - 10%：给领域的从业者带来灵感的启发
 - 1%：给行业所有的从业者都有启发，甚至带来方向性的影响

⚠：只有一小部分的文章值得复现，如顶会论文的best paper 或 oral；

⚠：经典的论文在你之前可能已经有很多人复现过了，作者也可能会给出官方的代码；

⚠：硕士、高年级的本科生、刚入门者这些垃圾可以不用读；但博士生为了在这个特别小的方向上能有所建树、保持敏锐的嗅觉，这些垃圾也要读，宁滥勿缺。

2. 论文来源

第一类：一手来源

- 会议的官网：cvpr、iccv、eccv、icml、nips（看best paper & oral & 获奖论文）
- arxiv.org (放着未发表的或者预览版的论文)，需要一个账号
- Google scholar; 订阅功能（自然语言处理、auto ml），订阅搜索结果，每日会自动搜索，关注最新论文

第二类：二手来源

- 野生收集：如github上的awesome list，可以订阅watch、star
- 可靠信息筛选：公开课（cs224n, cs224w, cs231n, 李宏毅老师机器学习公开课）的幻灯片里有论文列表一定要好好读；被人认可的个人来源、知乎大v、微博；项目源码的注释中的论文

第三类：三手来源

- 综述性论文

3. 如何读论文

场景：刚开完会议或正在开，手中有一次会议上所有被接收的论文或者随意按顺序拿出20篇

- **对20篇论文进行简单分类：**整理、分三类记录
- - **1. 论文标题和作者**
 - **2. 文章摘要**（用自己的话复述，用一句话或50字以内如何向别人推荐和介绍这篇论文，最好用英语；【这篇文章用了什么方法、在干什么】）

- 3.这篇文章要不要继续精读+当时作出决定的日期
- 精读环节：带着问题读【把自己想象成一个杠精去挑刺】（你开始提出的问题以及最关心的问题有没有在看完文章后得到解答）
 - 精读顺序：摘要（产生问题Q）-> 引言（上述问题是否被放大、有没有出现新问题、前面的问题有没有得到回答A）-> 再次确认要不要读下去 -> 总结部分，即conclusion、feature work（作者总结的文章重点以及未来的打算是否明确）-> 知识介绍（重点读每一段的第一句话；如果遇到图表，先仔细的看图表）-> 再次确定要不要继续读 -> 进入最耗时的核心算法部分 -> 实验部分（数据集、评价标准）-> 决定要不要复现、将自己的实验结果与论文进行比较
 - 相关工作：仔仔细细的看；这个方法是不是在某篇文章的基础上做的、与其他文章有什么关系、借鉴了哪些方法、以及该领域之前是用什么方法、有哪些数据集是被其他人做过实验的、其他人得到了什么效果；【related works】相关文章的题目 + 自己总结这篇文章与其他相关工作的关系或区别。

⚠ 时间：一个半小时；上述红色部分是笔记的记录内容；如果文章对于自己的难度较大，可以暂缓阅读，等把文章中提到的概念、需要的基础知识补一下再回过头来看。

最后在总结了所有的论文后，再附一张纸：总结多少文章都在关注什么方面的内容，有多少文章在针对某类技术、某个方向，大家都在研究什么。

⚠ 整个过程中要思考🤔：这个东西有没有价值、为什么大家都在做这些、你对他是什么看法。过几年之后，回过头来，可以看到随着时间研究内容发生的变化。

将这些样本数据，转化为自己的记忆是非常重要的。

4.文献管理工具

Zotero : <https://www.zotero.org/> (免费&跨平台)

Mendeley : <https://www.mendeley.com/> (免费&跨平台)

Endnote : <https://www.endnote.com>

Papers : <https://www.papersapp.com/>