

Paper课程01: How to Search & Read a Paper?

张老师

2020.06.17

Outline

- 关于NLP文献的选择和阅读
- Paper课程的内容和目标

Outline

- 关于NLP文献的选择和阅读
 - 论文检索
 - 文献选择
 - 阅读文献
- Paper课程的内容和目标

论文检索

- Google学术 （跟踪作者和论文）
- DBLP：<https://dblp.uni-trier.de> （整理会议、作者的全部工作）
- 微软学术：<https://academic.microsoft.com/home> （引用趋势）
- 中文检索数据库：
 - 万方
 - 维普
 - 知网

选择文献

- 面向特定主题的文献选择

- **Google Scholar :**

- 按作者搜索：author:"DM Blei", 可以搜索指定作者的相关论文；
 - 按发表期刊/会议搜索：source:"Nature", 可以搜索发表在指定期刊/会议的相关论文；
 - 按标题出现关键词搜索：allintitle:"question answering", 可以搜索在标题出现某些关键词的论文；
 - 搜索引擎常用的and、or和""均支持，其中""表示按引号中的字符串完整搜索。

- 面向某个领域的学习

- 中文知网（CNKI）中搜索"课题名称+综述"或在Google Scholar中搜索“课题名称 + survey / review / tutorial / 综述”来查找。
 - 相关领域会议/期刊tutorial

选择文献

- 面向知识更新的文献选择

- arXiv.org 上定期发布的论文；
- 相关国际顶级会议每年发表的论文集； ACL EMNLP COLING NAACL
 - <https://www.ccf.org.cn/xspj/rgzn/>
 - <https://dblp.uni-trier.de>
- 相关国际顶级期刊定期发表的论文；
 - Computational Linguistics, TACL
- 国际顶尖高校研究组或企业研究机构发布的新闻或学术报告；
- 科技媒体和社交媒体集中报道或讨论的学术成果；
 - 机器之心、雷锋网/AI科技评论、PaperWeekly、DeepTech、新智元为代表的技术媒体

选择文献

- 论文太多啦！！！！
- 被引用次数
- 会议或期刊的排名：可参考CCF列表
https://www.ccf.org.cn/Academic_Evaluation/By_category/
- 大牛作者或顶尖研究机构

选择文献

- NLP主要以自然语言文本为主要研究对象，与人工智能、机器学习、信息检索、数据挖掘、计算机视觉、知识工程等很多方向密切相关
 - 人工智能领域相关学术会议包括IJCAI和AAAI
 - 机器学习领域相关学术会议包括ICML, NIPS, ICLR
 - 信息检索和数据挖掘领域包括SIGIR、KDD、WWW、WSDM、CIKM、SIGMOD
- 计算机学会中文信息技术专委会组织的自然语言处理与中文计算会议（NLP&CC）是最近崛起的国内重要NLP/CL学术会议。
- 中文信息学会每年组织很多学术会议，例如全国计算语言学学术会议（CCL）、中国自然语言处理青年学者研讨会（YSSNLP）、全国信息检索学术会议（CCIR）、全国机器翻译研讨会（CWMT）等

走进一个领域

关键词、关键技术、重要论文列表、领域划分、领域大牛

- （相对成熟的领域） 综述和优秀的学位论文
- 领域扩充：
 - 关键词关联
 - 参考文献关联（相关工作）
- 会议扫墙，保持对领域的清晰认知
 - 扫墙的技巧：abstract和introduction

阅读文献

- 建议的阅读顺序：
- 题目 (1)
- 摘要abstract (2)
- 导论introduction (3)
- 实验结果results (4)
- 本文工作method (5)
- 相关工作related work (6) 、参考文献reference (6)
- 结论conclusion (7)
- 附录appendix (8)

阅读文献

- 快速阅读
 - Abstract
 - Introduction
- cs方向：Problem-Driven（对前人工作的优化）
- 理解论文核心思想，但是容易出现偏差

阅读文献

- 精读

- 理解论文基本原理: motivation, 所用的理论?
- 理解论文详细内容: 深入理解论文细节, 包括定义、假设和相关公式等
- ~~能够重现实验~~ (未必需要): 能否重现整个实验, 并得到相同结果?
- 组织讨论: 与其他人进行讨论, 可能有没有疏忽的地方? 或者理解不到位的地方?
- 设计更好的方案: 能不能更进一步, 设计更好的方法?

- **每个阶段可以对应一次或者多次阅读, 通过多次阅读来加深对论文的理解。不要指望一口气能够理解整篇论文。**

阅读文献

- Experiment is Important & Code is Everything
- 实验部分很重要 (作者对一些网络结构, 参数的设置为我们复现论文提供了重要的素材)
- 大部分论文中的公式依据我们现在的数学功底均可以理解, 很多公式直接与代码对应

Outline

- 关于NLP文献的选择和阅读
- **Paper课程的内容和目标**

Outline

- 关于NLP文献的选择和阅读
 - 论文检索
 - 文献选择
 - 阅读文献
- **Paper课程的内容和目标**

关于Paper课程

- 内容（结合正课的内容）
 - 整体把握
 - 局部重点
 - 其他延展（code, related work.....）
- 目标：
 - 领域相关知识、脉络的了解
 - Idea的启发
 - 时刻记住：不同于教材、做高等数学题，“每个字都认识，连起来读不懂”是正常的
 - 极为经典的文章：领域的奠基者

- THANKS & QUESTIONS?