刘知远：NLP 研究入门之道（二）走近 NLP 学术界

与老牌学科如物理学、化学等相比，计算机学科还非常年轻，学科体系长期处于剧烈变革之中。作为计算机应用的重要方向，人工智能和自然语言处理自然更不例外，与现实应用紧密相关，技术发展日新月异，常给人今是昨非之感。在这种情况下，传统学术期刊的那种投稿 1-2 年才能见刊的模式已经赶不上技术革新的速度，年度学术会议显然更符合计算机学科发展和交流的需求，可以看作是一种“小步快跑”的模式。阅读学术论文、参加学术会议是进入学术界、走进学术前沿的重要方式，在学术会议上，不仅可以集中听取最新的成果报告，还有讲习班（Tutorial）、工作坊（Workshop）、社交活动等形式，了解那些不会写到论文中的八卦与动态，结识学术大佬和朋友，走向学术人生巅峰。

国际学术组织、会议与论文

在计算机领域，国际上活跃着众多专业学术组织，吸收专业学者和学生作为会员，定期组织学术年会，报告学术论文，让学者们更方便地交流最新研究成果。这里以自然语言处理领域为例，介绍国际学术组织和学术会议的组织形式，以及国际学术论文的查找方式。

自然语言处理（Natural Language Processing，NLP）在很大程度上与计算语言学（Computational Linguistics，CL）重叠，与其他计算机学科类似，NLP/CL 领域有一个规模最大、最权威的国际专业学会，叫 The Association for Computational Linguistics（ACL，http://aclweb.org/ ）。ACL 学会主办了 NLP/CL 领域最权威的国际学术会议，即 ACL 年会。ACL 学会还在北美和欧洲设有分会，也定期召开年会，分别称为 NAACL 和 EACL。特别值得一提的是，2018 年 ACL 年会上宣布成立了亚洲分会 AACL，并定于 2020 年与亚洲另外一个著名国际会议 IJCNLP 合办第一届 AACL 分年会。

除了举办年会之外，ACL 学会下分设多个特殊兴趣小组（Special Interest Groups，SIGs），聚集了 NLP/CL 不同子领域的学者，性质类似一个大学校园的兴趣社团。其中比较有名的诸如 SIGDAT（Linguistic Data and Corpus-based Approaches to NLP）、SIGNLL（Natural Language Learning）等。这些 SIGs 也会自主组织相关主题的国际学术会议，其中最有名的应该是 SIGDAT 的 EMNLP（Conference on Empirical Methods on Natural Language Processing）和 SIGNLL 的 CoNLL（Conference on Natural Language Learning）。其中 EMNLP 发起于 1996 年，由于契合了近 20 年数据驱动的统计自然语言处理的发展脉动，因此受到广大学者的关注，也吸引了很多机器学习领域的学者参与。

国际上还有一个老牌 NLP/CL 学术组织 International Committee on Computational Linguistics，每两年组织一次学术年会 International Conference on Computational Linguistics（COLING），也是 NLP/CL 的重要学术会议。**NLP/CL 的高水平学术成果主要分布在 ACL、NAACL、EMNLP 和 COLING 等几个学术会议上。**

作为 NLP/CL 学者的一个重要福利是，ACL 学会网站用心建立和维护 ACL Anthology 页面（https://www.aclweb.org/anthology/），收录了 NLP/CL 领域绝大部分重要国际会议的论文全文并提供免费下载，甚至包括了其他学术组织主办的学术会议如 COLING、IJCNLP 等。新版 ACL Anthology 不仅支持基于 Google 的全文检索功能，还为每个学者建立了在这些会议上发表论文的主页，可谓一站在手，NLP 论文我有。

NLP/CL 领域也有自己的旗舰学术期刊，发表过很多经典学术论文，那就是 Computational Linguistics（http://www.mitpressjournals.org/loi/coli ），该期刊每期只有几篇文章，平均质量高于会议论文，时间允许的话值得及时追踪。由于审稿周期较长，近年来对学者投稿的吸引力下降，似乎论文质量也有所下滑。ACL 学会为了提高学术影响力，也创办了会刊 Transactions of ACL（TACL，http://www.transacl.org/ ），由于审稿周期与会议论文相当，并提供在各大学术会议上报告论文成果的机会，获得不少学者青睐，最近发表不少有影响力的工作，成长很快值得关注。值得一提的是，这两份期刊也都可以通过 ACL Anthology 开放获取。此外，也有一些与 NLP/CL 有关的期刊，如 ACM Transactions on Speech and Language Processing，ACM Transactions on Asian Language Information Processing，Journal of Quantitative Linguistics 等等。

根据 Google Scholar Metrics 2018 年发布的 NLP/CL 学术期刊和会议论文引用排名，ACL、EMNLP、NAACL、SemEval、TACL、LREC 位于前 6 位，基本反映了本领域学者的关注程度。其中 ACL、EMNLP、NAACL 的 H5-Index 和 H5-Median 明显高于其他会议和期刊，也是该领域每年参会人数最多的会议，可谓 NLP/CL 的三大顶级国际会议。另外，ACL 学会维护了一个 Wiki 页面（http://aclweb.org/aclwiki/ ），包含了大量 NLP/CL 的相关信息，如著名研究机构、历届会议录用率，等等，是居家必备之良品，值得深挖。

值得注意的是，虽然计算机领域学术会议论文的发表周期已经非常短，仍然不能满足最近深度学习等方向的迅猛发展。因此，越来越多学者选择绕过学术会议或期刊的审稿流程，直接通过 arXiv（http://arxiv.org/ ）等预印本平台在线发布论文。由于省去了同行评议的流程，这些最新学术成果得以更快地发布。但也由于缺少同行评议的意见和过滤，导致预印本平台上发布的论文质量良莠不齐，需要有较强的鉴别力，才能找到其中真正有价值的工作。毋庸置疑，arXiv 已经成为深度学习和自然语言处理最新进展的重要发布渠道，Yoshua Bengio 等著名学者及其团队的最新研究成果，往往先发布在 arXiv 上，然后再发表在相关顶级会议上。因此，arXiv 是了解大数据智能最新进展的重要信息渠道。

由于 arXiv 预印本客观上的确冲击了 NLP/CL 学术会议审稿的双盲规则（投稿作者和评阅人互相看不到对方身份），相关学者对通过 arXiv 率先发布成果看法不一，众说纷纭。从 2018 年开始，ACL、EMNLP、NAACL 等会议为了更好地执行双盲规则，对此提出了一种折中方案，将投稿截止时间前 1 个月也纳入匿名时段，即从投稿截止前 1 个月到稿件得到录用/拒稿通知，都不允许作者将具名论文发布到 arXiv 等预印本平台；对截稿前 1 个月以前发布到 arXiv 上的论文，也不允许在匿名时段再做更新或做媒体宣传。也就是说，从学术会议审稿公正性而言，并不鼓励将成果预先发布到 arXiv 预印本平台上。估计对这个问

题的争论还会持续，也许未来的确需要探索一种更好地兼顾高效与公平的学术论文发表机制，这是题外话就不再展开。

相关领域的国际学术会议与期刊

NLP/CL 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。例如，自然语言处理是人工智能的分支，而且人工智能的机器人、决策、知识表示等研究领域也与自然语言处理有交叉重叠；自然语言处理很多模型方法都来自机器学习的最新进展，自然语言处理也为机器学习提供独特的学习任务进行研究；信息检索关心的查询词、文档等也是自然语言文本，因此与自然语言处理关系密切；社会媒体中的用户生成内容很多为文本形式，是数据挖掘和自然语言处理共同关心的对象；计算机视觉和自然语言处理共同关注跨模态智能处理技术，如图像描述生成（Image/Video Captioning)等；知识和语言的天然关联性，也决定了知识工程与自然语言处理的交叉合作。这里主要介绍几个重点相关领域的国际学术会议与期刊。

人工智能领域相关学术会议包括 **IJCAI** 和 **AAAI**。AAAI 全称美国人工智能年会，IJCAI 全称人工智能国际联合大会。这两个会议方向非常广泛，涵盖机器人、知识、规划、自然语言处理、机器学习、计算机视觉等几乎所有 AI 子领域，是 AI 领域“奥运会”式的学术会议。近年来，由于 AI 领域备受社会各界关注，这两个会议的录用论文数也成倍增长。以 AAAI 2019 为例，投稿数猛增至 7000 多篇，最终录用 1150 篇，录用率降低至 16.2%。有些老师在社交媒体上如此评价，AAAI/IJCAI 更像花样齐全的“奥运会”，而 ACL/EMNLP/NAACL 更像专业领域的“锦标赛”，所以一般对专业领域任务的精细研究，更多发表在锦标赛式的专业会议上。由于知识表示等方向没有更权威的专门学术会议，所以更多发表在 AAAI/IJCAI 上。人工智能领域相关学术期刊包括 **Artificial Intelligence**、**Journal of AI Research**。

机器学习领域相关学术会议包括 **ICML**，**NIPS**，**ICLR**、**AISTATS** 等。其中 NIPS 全称是 Conference on Neural Information Processing Systems，由于最近这波 AI 浪潮就源自以神经网络技术为基础的深度学习，所以近年来备受关注，参会人数倍增，近几年会议注册页面刚开放就会被抢注一空。树大招风，2018 年由于 NIPS 缩写有性别歧视的意味，所以从 2019 年开始更名为了 NeurIPS。ICLR 是深度学习兴起后在 2013 年创立的年轻会议，采用的开放审稿模式，整个审稿过程的审稿意见、作者回复全部实时公开，也允许其他围观用户评论，面貌一新，关注者众，颇领一时风气之先。机器学习领域相关学术期刊主要包括 **Journal of Machine Learning Research**（**JMLR**）和 **Machine Learning**（**ML**）等。

信息检索和数据挖掘领域相关学术会议主要由美国计算机学会（ACM）主办，包括 **SIGIR**、**KDD**、**WWW**（从 2018 年开始更名为 The Web Conference）、**WSDM**。信息检索和数据挖掘领域相关学术期刊包括 **ACM TOIS**、**IEEE TKDE**、**ACM TKDD**、**ACM TIST** 等。其中 ACM TOIS 和 IEEE TKDE 历史比较悠久，地位卓然；ACM TKDD 则创立于 2007 年，ACM TIST 创立于 2010 年，均为新兴的著名期刊，特别是 ACM TIST 创刊时就邀请了 LibSVM 等有影响力的成果发表，现在 SCI 影响因子比较高。

中国计算机学会（CCF）制定了“中国计算机学会推荐国际学术会议和期刊目录”，基本公允地列出了每个领域的高水平期刊与会议。大家可以通过这个列表，迅速了解每个领域的主要期刊与学术会议。

国内学术组织、会议与论文

对很多学生（即使国外学生）而言，想参加 ACL、EMNLP、NAACL 等国际会议并非易事，由于注册费和差旅费很高，一般要有论文发表导师提供经费支持，而且长途跋涉也充满了签证申请、旅馆预订等不确定因素。作为学生，每年能出去成功且安心地参加一次国际会议，已然很不容易了。近年来，很多国内 NLP 学者已经可以持续发表高水平论文，进入国际一线研究行列，并与很多国际著名学者建立起密切的学术交流与合作。在他们的努力组织下，这些国内 NLP 学术会议的学术报告质量也有大幅提升，特别是特邀报告、讲习班、专题论坛等环节。需要说明的是，最近 AI 领域大火，国内很多机构都开始组织各类 AI 大会，其中很多特邀讲者不乏大牌学者。但为了强调学术导向，这里只聚焦那些以学术交流为主的纯学术会议。

与国际学术组织和会议相似，国内也有一家与 NLP/CL 相关的专业学术组织，中国中文信息学会（CIPS，http://www.cipsc.org.cn/ ），是国内最大的自然语言处理学术组织，最早由著名科学家钱伟长先生发起成立。通过学会的理事名单（http://www.cipsc.org.cn/lingdao.php ）基本可以了解国内从事 NLP/CL 的主要单位和学者。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，是国内 NLP/CL 学者进行学术交流的重要平台。尤其值得一提的是，YSSNLP 是专门面向国内 NLP/CL 青年学者的研讨交流会，采用邀请制参加，大家自愿报名在研讨会上报告学术前沿动态，是国内 NLP/CL 青年学者进行学术交流、建立学术合作的绝佳平台。2010 年的 COLING 和 2015 年的 ACL 在北京召开，均由中文信息学会负责组织工作，这在一定程度上反映了学会在国内 NLP/CL 领域的重要地位。此外，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。中文信息学会主编了一份历史悠久的《中文信息学报》，是国内该领域的重要学术期刊，发表过很多篇重量级论文。此外，国内著名的《计算机学报》、《软件学报》等期刊上也经常有 NLP/CL 论文发表，值得关注。

全国计算语言学大会（CCL）

CCL 是中国中文信息学会的旗舰会议，由 CIPS 的计算语言学专委会举办。CCL 从 1991 年开始每两年举办一次，从 2013 年开始每年举办一次，2018 年是第十七届。经过 20 余年的发展，是国内自然语言处理领域权威性最高口碑最好规模最大（2017 年注册人次超过 1 千）的学术会议，是国内 NLP 学者每年都会参加的盛会，现场交流氛围极佳。CCL 设置的讲习班、特邀报告、NLP 任务评测、前沿动态综述等环节，均有较大影响力，也是快速了解 NLP 前沿动态的绝佳方式。

其中，CCL 的特邀报告环节最具特色，CCL 程序委员会主席孙茂松教授每年都会大力邀请多学科相关重量级学者担纲。以 CCL 2017 为例，特邀讲者包括了中国工程院院士、西安交通大学郑南宁教授，清华大学社会科学学院院长彭凯平教授，中国香港科技大学计算机科学与工程学系系主任杨强教授，北京大学统计科学中心联席主任耿直教授，搜狗公司总裁王小川等，主题涵盖认知科学、心理学、机器学习、统计学等方向，议题与内容极具启发性。

全国知识图谱与语义计算大会（CCKS）

CCKS 由 CIPS 的语言与知识计算专委会举办，由国内两个相关会议合并而来，分别是中文知识图谱研讨会（CKGS）和中国语义互联网与 Web 科学大会（CSWS）。CCKS 是国内知识图谱、语义技术、链接数据等领域的核心会议，2017 年有 500 位学者注册参加。CCKS 设置的讲习班、工业论坛、评测竞赛、知识图谱顶会回顾、特邀报告等环节，具有较大影响力，是快速了解知识图谱等方向前沿动态的绝佳方式。

全国社会媒体处理大会（SMP）

SMP 由 CIPS 的社会媒体处理专委会举办，SMP 2018 是第七届，是国内聚焦社会媒体、面向社会计算和计算社会科学交叉学科的权威会议，SMP 2017 年有 800 多人次参加。SMP 也设置有讲习班、专题论坛、评测任务等环节。

其中，SMP 专题论坛非常活跃，以 SMP 2017 年为例，共设置了智能金融、计算社会学、情感分析、推荐系统、计算传播学、智能教育、表示学习及企业论坛等 8 个论坛，均有相关领域重量级学者担任讲者进行交流。

全国信息检索学术会议（CCIR）

CCIR 由 CIPS 和 CCF 联合主办，是中国信息检索领域最重要的盛会。会议除包含大会报告、论文报告、Poster 交流、评测活动外，还组织青年学者论坛、博士生指导论坛，以及面向热点研究问题的前沿讲习班等。大会也会邀请部分相关国际期刊、会议（如 TOIS、SIGIR、WWW、WSDM、CIKM）的中国作者交流论文。

全国机器翻译研讨会（CWMT）

CWMT 从 2005 年开始举办，2018 年是第 14 届，其中共组织过七次机器翻译评测，是国内最权威的机器翻译学术会议。除了传统的论文宣讲、特邀报告等环节，最近还设置了新人秀、产业论坛等环节，从事机器翻译研究与开发的同学不能错过。

自然语言处理青年学者研讨会（YSSNLP）

YSSNLP 是 CIPS 青年工作委员会的学术年会，其特色是采取邀请制，只允许青工委委员及其邀请的代表参加，每年约有 150 位青年学者参加，几乎囊括国内从事 NLP 研究的所有青年学者。青工委非常活跃，除了组织 YSSNLP 年会外，青工委还组织大量的国际顶级会议预讲会、学术沙龙等学术活动。

其中国际顶级会议预讲会是青工委的品牌活动之一，每年在 ACL、SIGIR、IJCAI、AAAI 等国际顶级会议正式召开之前，邀请国内有论文发表的学者介绍自己的论文工作。每次活动都吸引了大量来自学术界和工业界的现场和在线听众，极大促进了国内相关领域研究的发展以及研究者之间的交流。2018 年学术活动安排如下，欢迎大家关注并积极参与。

CIPS 暑期学校（CIPS Summer School）

这是 CIPS 的老牌学术活动，旨在面向青年学生进行前沿课题的教学与普及工作，带领同学迅速进入前沿。2018 年将是 CIPS 暑期学校的第 13 届。以 2016 年和 2017 年的暑期学校为例，均以深度学习技术在 NLP 中的应用开展教学，邀请国内一线青年教师和博士生担任讲者，系统深入地介绍深度学习的相关知识与动态。暑期学校每次持续 4 天课程，由于其较好的系统性和连续性，受到国内同学的广泛好评，近两年注册人数都超过场地容量。我个人担任了 2016 年暑期学校的讲者，以及 2017 年暑期学校的组织者，感觉这是非常好的系统学习 NLP 前沿动态的方式（虽然收费有点高）。

值得一提的是，从 2016 年起，CIPS 暑期学校被纳入到了 CIPS《前沿技术讲习班》编制，而 CIPS 组织的各大学术会议的讲习班也编入 CIPS《前沿技术讲习班》，由 CIPS 统一保证讲习班质量。

CCF 国际自然语言处理与中文计算会议（NLPCC）

NLPCC 由 CCF 中文信息技术专委会举办，NLPCC 2018 是第七届。NLPCC 按照国际会议模式组织，组织委员会注重吸纳国际学者，论文报告均用英文进行，是近年来国内崛起的重要 NLP 学术会议，2017 年参会人数超过 500 人，是在国内了解 NLP 前沿动态的又一个重要平台。值得一提的是，CCF 学科前沿讲习班（ADL）类似于 CIPS ATT，也是面向各类专题开展的讲习班，是 CCF 的老牌学术活动。NLPCC 每次都会附带一次面向 NLP 的 CCF ADL 讲习班，值得关注。

希望以上信息能够对初入 NLP 的青年同学有所助益。国内差旅成本较低，相信大部分导师会乐意支持学生参加学习，快速提高。最后想说，以上总结并非排名，仅为青年同学提供学习与交流的入口。而且限于个人所见，该总结难免挂一漏万，欢迎各种建议和意见，我会努力吸取改进。

结语

这篇介绍了自然语言处理领域国内外的主要学术组织、会议和论文，参加学术会议，阅读学术论文，是走近学术界、了解学术动态的主要方式，再辅以社交媒体和科技媒体，相信可以让同学比较及时地掌握自然语言处理科研动态。