

保持高效论文写作的10个原则！

机器学习算法与自然语言处理 2020-11-24

公众号关注 “[ML_NLP](#)”
设为“[星标](#)”，重磅干货，第一时间送达！



对于科研人员来说，论文的写作与阅读一项核心技能。实际上，能否发表论文已经成为评判科研人员能力的标准，并预测着他们未来的成功之路。

在论文的产生与消费中，有多方人士参与其中，然而，每一方都有他们自己的目的和偏好：编辑希望你的论文意义重大；审稿人需要判断文章的结论是否证据确凿；读者则希望能快速理解主旨，以决定是否深入研读文章细节。

对于作者自身而言，则是一方面尽可能地向广泛的受众传递文章的重要贡献，另一方面说服专家认同研究成果。**为了实现所有的目标，你需要将文章在多个方面都组织得尽善尽美——从句子、段落、章节到相关材料。**



对于更加广泛的科学事业来说，清楚明确的交流也不可或缺，因为“概念迁移”成为了交叉学科发展的阻碍。特别是在生物学科与其他学科的交叉过程中，产生了庞大的、高度互联的次级学科网络。随着研究人员愈加专业化，强化概念间的联系变得愈加重要，但也愈发困难。跨学科的交流仅能建立在文章具有可读、可信、可记忆的基础之上。

因此，我们给出了十种简单的规则来规划论文结构。

第一条规则：文章要有一个明确的中心思想，并要出现在文章的标题中。

第二条规则：为那些不清楚你工作的有血有肉的人们写作。

第三条规则：坚持背景-内容-结论（CCC）的结构组合。

第四条规则：优化你的逻辑流：避免曲折，使用平行叙述。

第五条规则：在摘要中讲述完整故事。

第六条规则：在介绍中解释清楚为什么这篇论文至关重要。

第七条规则：用多个逻辑相关的陈述句（可辅以图表）给出支持论文中心思想的结果

第八条规则：讨论空缺如何被填补，说明诠释中的限制及同领域的相关性

第九条规则：将时间分配到重要的部分：标题、摘要、图表和提纲

第十条规则：获取反馈，然后简化、重新使用、再次构造这个故事

前四条规则可以应用于一篇论文的所有部分，甚至是其他形式的交流，如会谈和海报等。中间四条规则涉及论文每个主要部分的主要目标。最后两条规则为有效构建论文原稿提供了过程启发式的指导。接下来就随我们一起，了解论文写作的奥义。

01

文章要有一个明确的中心思想 并要出现在文章的标题中

写作即是交流。因此，读者的体验至关重要，所有的写作都应这一目标而服务。当你写作时应当时刻将读者铭记在心。下面四条规则将会帮助你避免读者的流失。如果你的读者在一年之后，仍能向他们的同事描述你文章中的中心思想，那表明你的沟通工作卓有成效。

虽然我们需要在文章中描述大量的创新点，但是着眼于一条单一的信息更加重要。同时关注过多的贡献会使得文章缺乏说服力，并变得难以记忆。

标题是文章中最重要要素——想一想你所阅读的标题与论文数量的比例吧。**标题作为读者第一眼看到的东西，直接决定了他们是否会继续花时间阅读摘要。**标题不仅传递文章的中心思想，同时还不断地提醒你，要在文章内容中处处传达这个思想。毕竟，科学就是从复杂数据中提取出抽象的简单原理的过程。因此，标题应该浓缩论文的核心思想。提早考虑给论文一个标题，边写边琢磨如何改得更好——这不仅有助于你的写作，同样有助于实验设计和理论发展。

第一条规则成为对论文优化中最难的一条：**你将会为平衡多方面的结果，最终合成一条中心思想而努力。**举个例子来说，一篇关于技术的文章可能会同时描述使用该技术产生的生物学结果，而统一这两个方面的桥梁，便是清楚地描述新技术如何应用于新生物学之中。

02

为那些不清楚你工作的有血有肉的人们写作

你是世界上最了解你所做工作的专家，因此你也成为了最不合格的读者，你无法从不了解你工作的读者角度去评判你文章的人。在这一困境中产生的最为多数的写作错误是从一个设计

者的角度进行思考 --- 决定你想在每一个元素上对人们产生的影响，然后努力实现这一目标。

试着站在一个“懵懂的”读者角度去思考这一论文，让读者更轻松的快速且不费力的抓住信息。

人类心理学的广阔知识同样在论文写作中十分有用。举个例子，人类拥有工作记忆的限制：他们只能记住少量项目，同时对起始和结束的部分印象深于对中间部分的记忆。利用你自己的心理学知识，以“懵懂”读者的视角去引导写作的进程。在下面的原则 10 中，我们将对在写作过程中从“懵懂”的试验性读者那里得到反馈从而获得的效益展开讲述。

03

坚持背景 - 内容 - 结论 (CCC) 的结构组合

大多数广为传播的故事都有着类似的结构。开头为故事背景的设置，主体部分（文章内容）展开故事，向结尾推进，最后得出文中问题的结论。这样的结构可以减少读者产生疑惑的机率，他们不再会问：“为什么会这样”、“所以呢？”

讲故事的方法多种多样，C-C-C 结构作为其中一种，更适用于有耐心、决定花时间理解文章的读者，但这种结构对没有耐心的读者并不友善。好在当代科学文章的结构解决了这样的问题，标题与摘要使读者可以快速了解文章内容。因此，看完引言的读者很可能被吸引，并有着足够的耐心继续阅读文章。而且，过于注重“内容优先”的科学论文还会产生另一个风险：读者在阅读过程中会忽视一些段落，这些段落可能是巩固文章可信度的重要部分，从而使他们质疑你的文章。因此，我们建议将 C-C-C 结构作为科学文献的默认结构。

C-C-C 结构在多个尺度上定义了文章的结构。从整篇文章来说，引言设定了背景，实验结果是文章内容，讨论得出结论。从段落来说，首句定义了主题或背景，主体部分提供引发读者思考的新颖观点，末句则为读者提供需要记住的结论。

不按照 C-C-C 结构写作的论文往往难以阅读。在科研生涯中，我们在科研内容上花费了过多时间，从而忽略了其他方面。我们按照科研的时间顺序来记录与组织文章，殊不知对于读者而言，他们并不关注这些细节，他们仅仅对最终的结论与支撑结论的逻辑感兴趣。

04

优化你的逻辑流：避免曲折，使用平行叙述

文中只有中心思想应该多次被提及。此外，为了减少主题改变的次数，每一个主题仅需提及一次即可。相关的句子与段落应该连贯，避免被不相关的材料打断。类似的观点应该紧密相连，例如应当相信某件事的两个原因。

同样地，对于连续的段落或句子，平行的信息应该使用平行的形式来阐述。平行方法使读者更加熟悉结构，使文章更易阅读。举个例子来说，如果有三个独立的原因让我们更偏向某一解释，我们应该使用相同的句法表达，这样的表达对于读者而言清晰明了，使他们更加专注于内容。需要避免使用不同的词语指代同样的概念，因为这样做会使读者对不同词语的意义产生疑惑。

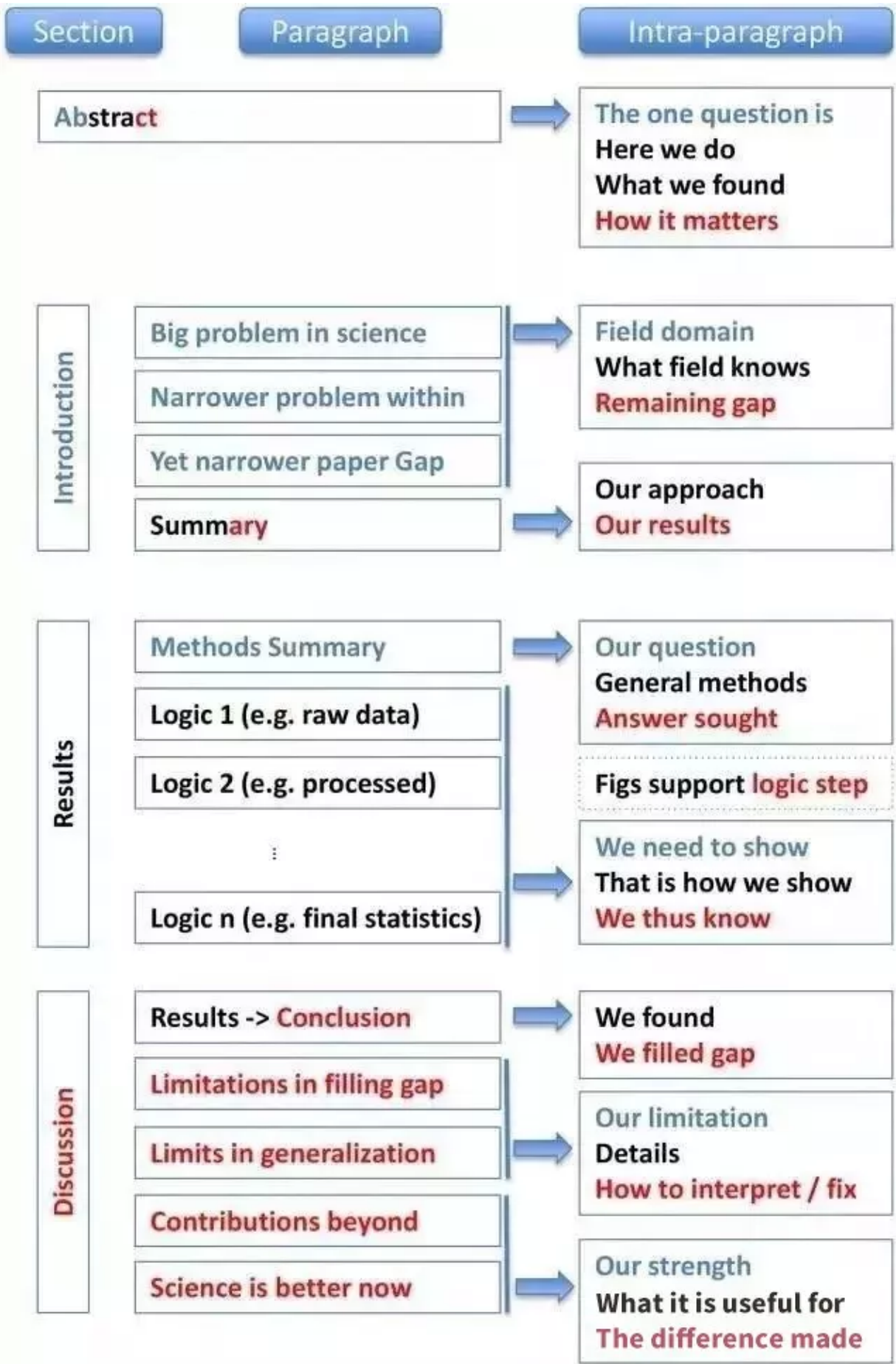


图 1. 论文三个重要部分的结构总结。Note：摘要是最特别的，它包含所有三个要素（原因、内容和结论），所以包含了所有三种颜色。

在摘要中讲述完整故事

论文的各个部分 --- 摘要、介绍、结果和讨论 --- 有着不同的主题，因此整体结构和段落结构稍有不同。对于大多数读者们来说，摘要是论文中唯一会被阅读的部分。这意味着摘要中必须有效呈现论文的所有信息。为实现这一目标，摘要的结构必定是高度保守的。每一个 C-C 元素中均有两个部分，细节如下。

背景必须解释清楚文章即将填补的空缺。第一句话通过介绍该研究所处的广阔领域对读者进行引导。随后这一背景缩小到该研究所要回答的待解决问题上。通过表述“文献中缺失部分”（例如，特定的空缺）和“为何这很重要”（例如，这一特定空缺和论文起始的广阔背景之间的联系），一个成功的背景章节能够为从现有技术水平中分辨出论文所表述的贡献打好基础。

这里我们所说的内容首先描述的是能够让你回答这一空缺 / 问题的新颖方法或者途径。随后你呈现出“血肉”--- 获得这些结果的执行概要。避免使用含有高度科技含义的词汇（术语 --- 这些词汇会让你失去读者们），尽可能少使用含有精妙含义的词汇。

最后，结论会通过阐明结果从而回答在背景章节最终部分提出的问题。随后，尤其对多数拥有广阔读者群的“常规”科学杂志来说，通常结论会有第二部分，突出讲述该结论如何推动广阔领域继续前进（例如，“更广阔的重要意义”）。

这一结构能帮助你避免最为普遍的摘要错误：在读者们准备好理解它们之前便谈论结果。优秀的摘要通常多次反复进行提纯细化从而确保结果能够完整填补空缺部分，就如同锁和钥匙一般。这种宽阔 - 狭窄 - 宽阔的架构方式能够让你同广大读者们（通过广度）进行交流，同时保持你结论声明的可信性（这通常是基于一套有限 / 狭隘的结果）。

06

在介绍中解释清楚为什么这篇论文至关重要

介绍突出体现了存在于现有知识或者方法中的空缺和它为何如此重要。这通常是由一系列渐进的更为精确的段落来完成，这些段落最终会以对文献中缺失部分的清晰阐述而结束，另一段落会紧随其后，对文章如何填补空缺进行总结。

举一个空缺渐进的例子，第一段落也许解释了为什么细胞分化是一个重要的课题以及这一领域中还未解决到底是什么触发了它（领域空缺）。第二段落则阐明了某一特定细胞类型如星形胶质细胞分化过程中的未知问题（子领域空缺）。第三段落也许提供线索，一个特定基因

或许驱动星形胶质细胞的分化，随后声明该假设尚未得到验证（将要进行填补的子领域空缺）。这种空缺叙述引起读者对这篇论文想要传递思想的期待。

每一个介绍段落的结构（除去最后一段）均会促进空缺发展的这一目标。每一段落首先会引导读者走向主题（背景），随后解释相关文献中所知所想（内容），然后引导进入至关重要的“未知”（结论），这一部分使得文章在相关领域十分重要。

在这一过程中，通常会有很多线索提供隐藏在空缺后的秘密，指向未经验证的假设或者文章中未开发的方法，其中我们了解到为什么填补空缺是具有前途的。介绍中不应含有超越文章动机范围外的文献综述。这种“空缺专注”的结构使得经验丰富的读者们更为容易地对一篇论文的潜在重要性进行判断 --- 他们仅需要评估声明空缺的重要性。这些空缺描述即该篇文章的贡献将会被以如下条件进行衡量：以文章空缺定义来设定的兴奋性、期望值和高质量背景。

介绍的最后一段则十分特殊 --- 它简单总结了用于填补上述空缺的结果。它同摘要有所不同，不需要显示背景（已在上述内容中描述过），但需要稍微更准确详细地阐述结果，如果可能的话可以简短地预先描述论文的结论。

07

用多个逻辑相关的陈述句

（可辅以图表）给出支持论文中心思想的结果

结果章节需要说服读者相信中心声明是有数据和逻辑来支持的。每一个科研论据都有它自己独特的逻辑结构，从而决定要素该如何呈现的顺序。举例说明，一篇文章也许建立起某个假说，核实在研究的体系中测量的某个方法是否有效，随后运用这种测量手段反证该假说。另一种选择是一篇论文也许建立了多个可选性（且相互排斥的）假说，除去其中一个能够提供剩余解释的证据外，对其他所有都进行了驳斥。论据的组织结构应当含有整体逻辑所需的对照和方法。

在文章准备的提纲阶段（见原则 9），先草拟出这些结果如何支持你的主张的逻辑结构，然后将它们转变成陈述性语句，用于结果（results）章节的子章节标题。大多数学术期刊允许该类型格式。即使你选择的期刊不允许，这些标题在写论文的时候也很有帮助，可以作为段落的开头介绍，或者在提交之前删除。这种清晰的逻辑结果使读者更容易跟上论文的思路。

图、图的标题和图注非常重要，因为他们展示了对论文论点最客观的支持数据。此外，为了节省时间，有些读者会略过摘要，直接选择看图。因此，图的标题应该传达分析结论，图注

应该解释结论如何得出。理想情况下，这些图片应该生动的讲述整个故事，而不需要阅读图例或者文字。制图本身也是一门艺术。Edward Tufte的书介绍了学习这门技艺的黄金法则。事实上，有些读者仅阅读摘要和图片。理想情况下，这些图片应该生动的讲述整个故事，而不需要阅读图例或者文字。

【results】部分的第一个段落特别之处在于，因为它通常概括解决引言所述问题的所有方法，以及论文提出的重要创新方法。大部分读者不读【methods】部分，因此这一段可以帮助读者大致了解论文使用的方法。

【results】章节后续的每一段都应一两句介绍这一段要解答的问题，比如“”开头可以用一两句话提出这个段落所要回答的问题。例如，“为证明这其中没有任何人工干预”，“什么是我们测量的重测信度？”或者“我们进一步检测通过 L 类型 Ca^{2+} 通道的 Ca^{2+} 流是否参与其中”。段落的中间部分展示问题涉及的数据和逻辑，然后以问题答案结尾。举例来说，它也许推断出没有任何可能的人工干预被检测到。这样的段落结构可以让有经验的读者对论文进行事实核查。每一段的最后一句是总结句，这样很方便就可以找出哪一个结论值得质疑，以及检查每一个的逻辑。每段的结果都有一个逻辑陈述，后续的段落则依赖于先前段落的逻辑结论，就如同数学文献中的定理一样。

08

讨论空缺如何被填补

说明诠释中的限制及同领域的相关性

讨论章节解释了为什么结果能够填补介绍中提出确认的空缺部分，对诠释说明提供预警，以及描述该篇论文如何通过开创新的机遇从而推动领域发展。这通常以如下方法完成：对结果进行总结概括，讨论限制点，随后揭示这一中心贡献如何促进未来的发展历程。讨论章节第一段落的特别之处在于它通常大致总结了结果章节中的重大发现。有些读者跳过结果章节的大部分内容，因此这一段至少能够带给他们对于上述章节的大致情况。

讨论章节接下来的每一个段落都以描述该文的某个缺点或者优点作为开端，随后通过广泛链接到相关文献从而对优势或者弱点进行衡量。最后，它将作者对于结果贡献的感知进行总结，并探讨进一步研究的可能性途径。

举例来说，第一段落也许对结果进行总结，专注于它们的含义；第二至第四段落也许讲述了潜在的弱势，该文章如何缓和这些情况或者未来的实验能够如何处理这些弱点。第五段落以及普遍存在的第六段落也许以该论文如何推动整个领域前进而告终。如此循序渐进，读者们最终学会将文章的结论代入到正确的情境中去。

为撰写一篇优质论文，作者们可以采用具有帮助性的流程和习惯。对于一篇论文的影响力而言，某些方面比其他方面更为重要，因此建议你应该将时间投资到影响最大的问题上。此外，使用从同事那里获得反馈进行反复改进可以让作者从整体水平上提高文章从而撰写出强有力的原稿。选择正确的流程能够让文章撰写更为轻松和有效。

09

将时间分配到重要的部分：

标题、摘要、图表和提纲

科技文章的中心逻辑至关重要。它同样是连接论文研究试验阶段和论文写作阶段的桥梁，因此将文章的逻辑结构同正在进行的实验保持一致十分重要。由此来看，将正在进行的实验工作的逻辑正式形成某种不断展开的文档十分有用，最终它将成为论文提纲的蓝图。

你还应当按照各章节的重要程度分配时间。看题目、摘要、图的读者远远多于查看论文其他内容的读者，【methods】部分是阅读者最少的章节。所以，根据重要程度合理分配时间。

我们可以在写作之前先进行规划，从而高校利用各章节所花费的时间。写一个大纲。我们喜欢为一个计划要写的段落写一个不太正式的句子。这通常有利于开始进行该段落的写作，而这些句子有可能成为节标题（section header）。文章有一个整体框架，因此每一段都应该担任推进文章的特定角色。这个角色最好在大纲阶段就进行详细检查和确认，减少不适合文章的段落上浪费时间。

10

获得反馈，对整个 论文进行精简、再利用和再生

我们可以把写作看作一个优化问题，你需要同时优化故事、大纲和所有句子。因此，不过分依赖自己原本的写作内容非常重要。在很多情况下，删去整段文字和重写是比在原有的内容上进行编辑更快速的输出优质内容的方式。

很多信号提醒大家必须修改论文手稿（见表1）。例如，如果你无法在几分钟内向同事描述论文的大纲，那么很明显读者也无法厘清大纲。你必须进一步精炼论文内容。找出这些与优秀写作违背之处进行改正，可以帮助你全面提高文章质量。

Rule	Sign it is violated
1: Focus on one big idea	Readers cannot give 1-sentence summary.
2: Write for naive humans	Readers do not “get” the paper.
3: Use context, content, conclusion structure	Readers ask why something matters or what it means.
4: Optimize logical flow	Readers stumble on a small section of the text.
5: Abstract: Compact summary of paper	Readers cannot give the “elevator pitch” of your work after reading it.
6: Introduction: Why the paper matters	Readers show little interest in the paper.
7: Results: Why the conclusion is justified	Readers do not agree with your conclusion.
8: Discussion: Preempt criticism, give future impact	Readers are left with unanswered criticisms and/or questions on their mind.
9: Allocate time wisely	Readers struggle to understand your central contribution despite your having worked hard.
10: Iterate the story	The paper’s contribution is rejected by test readers, editors, or reviewers.

<https://doi.org/10.1371/journal.pcbi.1005619.t001>

表 1. 十条规则以及如何辨别文章违反了这些规则。

完成一篇成功的论文需要多方面人力的投入。测试读者是必须的，他们可以确保整篇故事起作用，同时他们也能够给出具有价值的意见，如论文的那一部分看上去进程过快或者过慢，此外他们也能够明确在什么时候最好需要重头再来和重述整个故事。审稿人同样十分有用。

非科技反馈和冷漠的评审结果通常说明评审没有【get】整个文章脉络。非常具体的反馈通常指出某一段逻辑不足。用积极的方式接受反馈非常重要。来自他人的反馈非常必要，因此找几个可以在这方面提供帮助的同时是提高论文质量的基本方法。为了维持好关系，不要忘了也阅读他们的文章作为回报。

本文的重心在于论文的结构，为此不得不省去了很多写作的细节问题，包括选词、语法、创造过程和合作。关于如何写作的文章可能永远都不完备，还有很多文章介绍如何进行科学论文写作。

个人风格通常使作者偏离死板、保守的文章结构，阅读创造性地改变规则的论文也是一种乐趣。但是，和生活中的很多事情一样，完全掌握标准规则是成功改变规则的前提。遵循上述原则，科学家才能够将知识传递给更多的读者，连接不同的学科领域，更高效地实现整合科学。

文章来源：学术出版社

下载1：四件套

在机器学习算法与自然语言处理公众号后台回复 **“四件套”**，

即可获得学习TensorFlow，Pytorch，机器学习，深度学习四件套！



下载2：仓库地址共享

在机器学习算法与自然语言处理公众号后台回复 **“代码”**，

即可获得195篇NAACL+295篇ACL2019有代码开源的论文。开源地址如下：

<https://github.com/yizhen20133868/NLP-Conferences-Code>

重磅！机器学习算法与自然语言处理交流群已正式成立！

群内有大量资源，欢迎大家进群学习！

额外赠送福利资源！深度学习与神经网络，pytorch官方中文教程，利用Python进行数据分析，机器学习学习笔记，pandas官方文档中文版，effective java（中文版）等20项福利资源



获取方式：进入群后点开群公告即可领取下载链接

注意：请大家添加时修改备注为 [学校/公司 + 姓名 + 方向]

例如 —— 哈工大+张三+对话系统。

号主，微商请自觉绕道。谢谢！



推荐阅读：

工业界求解NER问题的12条黄金法则

三步搞定机器学习核心：矩阵求导

神经网络中的蒸馏技术，从Softmax开始说起

喜欢此内容的人还喜欢

全世界都在努力做科研，只有我们在玩命的朝SCI灌水

机器学习算法与自然语言处理

轰鸣、弹射！“车神”养成进行时

联勤集结号

又一家米其林餐厅倒闭了，外面的世界还会好吗？

造洋饭书