

科研论文写作指导

余深宝^{*†}

[†] 自动化系, 厦门大学

2024 年 1 月 25 日

写在前面

科研论文写作是一个对我们现阶段的科研工作系统进行总结和逻辑重组的过程。在该过程中，我们既能够训练逻辑思维、锻炼思考能力，还能够举一反三地优化处事规则和工作习惯。一个好的科研论文不仅能够帮助自己厘清工作的来龙去脉，还有助于传播学术思想。因此，如何写好一篇科研论文是一个科研工作者的必备技能。但是，它也不是一蹴而就的，需要我们在每一次地科研训练中不断地打磨以便贴合自己的科研习惯。我写该指导的目的是对我在四年的博士生涯中与许许多多的前辈和同侪们进行交流后的经验总结；以及对自己所撰写的、导师们帮忙批改的、还有帮助实验室的师弟师妹们所修改的论文的反思；当然，更多的是对经常由于各种各样的原因而被拒稿的复盘。作为一个年轻识浅的科研工作者，这篇指导也有许多个人偏见，但我希望它能够给后辈们带来一些启发，少走一些弯路。

在这里，我会先详细介绍科研论文的主要组成部分（第1节），接着总结写好科研论文的几个技巧（第2节），然后谈一谈我在写作过程中的心得体会（第3节），最后列出自查论文的顺序和注意事项（第4节）。

1 科研论文写作框架

科研论文¹的写作框架主要包含：(1) 题目 (title)、(2) 摘要 (abstract)、(3) 引言 (introduction)、(4) 相关工作 (related work)、(5) 背景知识 (preliminary or background)、(6) 自己所提出的方法（模型和算法）论述 (the proposed model or algorithm)、(7) 实验 (experiments)、(8) 总结与展望 (conclusion and future work)、(9) 致谢 (acknowledgement)、(10) 参考文献 (reference)。其中，(1)、(2)、(3)、及 (8) 以简短的文笔对整篇论文进行“提纲挈领”；而 (4)、(5)、(6)、及 (7) 则是对之“细针密缕”。下面将对各个部分逐一进行写作要点的介绍。

1.1 题目

题目是论文的灵魂，一个好的题目会让审稿人和读者快速记住你的文章，并且先入为主地提高印象分。好多文章的工作做得很漂亮、内容写得优美、实验做得充实，却因为一个不起眼的题目得不到

^{*}yushenbao@outlook.com / yushenbao@stu.xmu.edu.cn

¹如无特殊强调，这里指的是以英语论文写作。

赏识。更要命的是，我曾经审过一篇稿子，它的标题直接误导了我对文章的定位，结果在审稿的过程中埋下了一直想要拒稿的冲动。另外，[陈怡然](#)老师在他的一篇博客²中说道：“在评审一篇文章的过程中，题目会被至少提及 3 次。那么如果两篇文章的质量差不多，一篇的题目平平无奇，另一篇的题目博人眼球，试问哪篇文章评审人会先读、精读、提出更有建设性的审稿意见？”因此，请不要吝啬时间为你的论文题目取个好名字。

对于如何起一个“好”的题目，说实话，目前我没有什么建设性的意见。我想只能通过多读文章、多积累一些漂亮的题目形式作为启发，例如：

- [A Penny for Your Thoughts?](#) the Value of Information in Recommendation Systems.
- Estimating Student Proficiency: [Deep Learning Is Not the Panacea](#).
- [“Why Should I Trust You?”](#) Explaining the Predictions of Any Classifier.
- [Causal Parrots](#): Large Language Models May Talk Causality But Are Not Causal.
- [DAGs with NO TEARS](#): Continuous Optimization for Structure Learning.

1.2 摘要（最后写）

摘要是对自己工作的总结，它是引言的缩写版，[但更突出论文的贡献](#)。它一般是一段 150–200 词的文字，主要讲清楚你的科研任务是什么、提出了什么方法（千万不要提细节）、以及有什么创新点。我们大致可以从以下三个方面进行描述：

1. 先介绍所研究的问题是什么，接着介绍该科研问题的重要性；
2. 然后介绍现有的解决方案及存在的不足之处（这个问题是与本文提出的方法相关的）；
3. 紧接着指出自己提出了一个什么方法来解决这个不足，同时详细介绍自己的方法：采用了什么样的模型，解决了什么不足，得到了什么样的结果。

1.3 引言（最先写）

良好的开端是成功的一半。引言的写作是一门“讲故事”的艺术，它最好的样子应当是[看起来不可或缺，做起来思路清奇、读起来如沐春风](#)。

1.3.1 引言的写法

如果不知道怎么写，可以用“六句扩充法”，写完这六句话之后，再去扩展即可：

1. **研究问题的背景和意义**。以社团检测任务为例子，在该任务中，我们需要介绍社交网络中社团的重要性的研究意义，包括社团是什么、社团检测是什么、并列举其应用场景。

²<http://www.tcse.cn/wsdou/advice/deadline-chen.pdf>

2. **研究现状以及主要问题.** 这个模块区别于相关工作 (related work), 它主要给读者一些研究背景的铺垫。如可从宏观的角度介绍解决该问题的几类主流方法, 包括其优缺点的分析和存在的不足。这个不足应立足于文章的创新点, 即相对于其它方法的改进点。
3. **解决问题的挑战.** 说清楚要解决的问题的难点在哪里。如果没说好甚至不说, 大家会觉得你做这个问题太简单了, 没什么难度, 那么文章自然而然地就没什么价值了。
4. **自己方法的解决思路.** 不要沉浸在自己的世界里, 把思路说得云里雾里, 让人摸不着头脑。
5. **自己方法的技术方案.** 不要面面俱到介绍所有的细节 (包括如何调参等)。我们需要写清楚最重要的部分, 有些辅助组件甚至可以不提。另外, 如果能画图就尽量画图, 因为一图胜千言。
6. **写总结、强调贡献点.** 创新点一般写三个, 包括但不限于“因为要解决什么问题而提出了什么模型”、“提出了什么算法”、“达到了什么效果”。

1.3.2 引言写作的常见错误

- **蓄意科研.** 因为模型 A 好用, 所以我用 A 做某任务。
- **暴力科研.** 因为任务 B 没人做, 所以我做了。
- **缺乏对比.** 之前的人做了什么工作, 我做了什么工作, 我的好。
- **盲目自夸.** 这是审稿人最反感的做法。例如说之前的工作这也不行那也不行, 就我的方法行 (千万不要说 “The deficiency of Wang’s approach is ...” 但可以委婉地说 “However, the mechanism has not been fully understood”、“None of the other phases have been examined in detail”)。

1.4 相关工作

该部分的段落数自行调整, 但每段不宜太长 (会议要短写, 期刊要长写), 且使用文字性描述即可 (一般不涉及公式)。以下是几点建议:

- **分类整理.** 涉及的主要任务有哪些、用到的数据有哪些、算法有哪些、模型有哪些等。
- **阐述各工作的区别与联系.** 写相关工作时很忌讳无脑地简单罗列, 如 A 做了一些什么, B 做了一些什么, C 又做了一些什么 (或是张三 proposed 什么, 李四 proposed 什么, 王五 proposed 什么)。
- **时态.** (1) 介绍别人的工作时采用一般过去时。如 “Scheirer *et al.* [6] employed a support vector machine for known class identification ...”; (2) 陈述某个方法时用一般现在时。如 “In contrast, DCMs classify students into multiple categorical attributes with mastery/non-mastery states aided by the Q-matrix.”。

可推荐的一种写作思路为: 以讲故事的形式介绍待解决的问题的已有方法 (包括 A 提出了什么方法, 简要介绍该方法的思想 and 存在的缺点; B 又提出了怎样改进的措施, 且还存在一些什么问题; C 又是如何改进的; 直到引出自己的方法)。

1.5 背景知识

以**文字和公式相结合的形式**介绍与自己所提出的方法直接相关的模型。例如，当自己的方法是某几个模型的融合时，那么在此处对它们进行详细介绍。注意介绍的公式是自己方法中会用到的，通用的模型（如 k-means 聚类算法）就不要太详细阐述了或者直接略过，除非自己真的没得写。

这里要特别注意数学符号的使用：(1) 要符合人的直觉，要好看，要好记，要成体系；(2) 活用上下标，但尽量不要用上下角标都存在的符号。关于数学符号的规范可以参考 <http://www.deeplearningbook.org/contents/notation.html>。

1.6 所提方法的论述

介绍自己提出的方法（**只描述自己提出的方法，不要讲不相关的东西**），段落数自行调整。这里给出两种基本的写作思路：

1. **“总-分”式**. 这个是我比较喜欢的写法。当模型有若干个部分时候，先从模型概述讲起以便做一个宏观概括，再分别介绍每个部分。
2. **“总-基础-增强”式**. 从基本模型开始写起，再在模型上做扩展。

另外，我们可以在该部分的末尾多作一些讨论，为方法“添彩”。讨论的内容主要针对模型/算法的分析，如：

- 正确性的证明；
- 模型的（时间）复杂度；
- 完整的算法（图）流程；
- 与先前工作的区别（突出创新性）；
- 与先前工作的联系（增强泛化性）；

1.7 实验

大致分为四个方面，具体段落数自行调整：

1. **数据集和评价指标**. (1) 数据怎么来的（表格内数字靠右，用逗号表示法）。一般写作思路为：原始的数据集 → 重要的预处理步骤 → 处理后的数据集。(2) 新任务的评价指标要给出详细说明，老任务的评价指标可以简写，以便压缩空间。
2. **对比方法**. 如果对比方法较多，最好用表格总结一下，让读者有一个整体的了解。
3. **主要实验结果**. 主要介绍与其他对比方法的比较情况。注意两点：(1) 存在异常的结果（如效果比对比方法差等）要解释分析；(2) 最好加一些统计性显著检验，确保提升是有效的。
4. **其他实验结果**. 如案例分析、参数敏感性分析等。

1.8 总结与展望

该部分一般为两段：

- 总结文章的工作：提出了什么方法、解决了什么不足、取得了什么结果。它一般与摘要和引言中的某些句子所表达的意思是类似的，因此要适当地去改变一下句式。
- 展望一下现在的研究工作还能朝着哪个方向继续进行改进，特别是想一想目前的方法有哪些潜在的问题，这些问题又有哪些是以目前的条件没有办法做的，做到“堵漏”。这样做既能够让人觉得我们对待解决的问题和所提出的模型做了全面调查和深入分析，也可以先发制人地防止审稿人提问。

1.9 参考文献

注意事项：

- 近 5 年的参考文献占全部文献总数 1/3；
- 信息要完整：作者、标题、期刊/会议名称、页码、年份；
- 注意文献的人名、文章标题、年份要一致。有些文章会预发表，正式发表时可能会有所改动，这时需要写最新的参考文献版本；
- 每个参考文献的人名格式要一致。如果人名要缩写（如 N. Ailon），那么都要缩写，不要有些写完整的人名，有些又是缩写；
- 每个参考文献的标题大小写要保持一致。如只有第一个单词的首字母大写还是所有单词的首字母都大写。

2 科研论文写作技巧

2.1 技巧一：写作需要逻辑性

科研论文的灵魂是**逻辑**，而逻辑又体现在 (1) 整篇文章的来龙去脉清晰可见；(2) 文章内各个部分，以及每个部分内句与句之间的起承转合自然无暇。

针对第一点，在构思文章时，看看是否包含了“3W2H”：

- 做了什么 (contribution)? → What
- 为什么做 (motivation)? → Why
- 怎么做的 (model and algorithm)? → How
- 效果怎样 (experiment and discussion)? → How much
- 还能做啥 (conclusion and future work)? → What then

其中 3 个 W 很重要，它表达了整篇文章的核心内容。我建议可以根据“3W2H”法则对自己的研究工作先拟一个草稿，方便根据之后出现的新想法进行调整。

针对第二点，用五个字概括就是“前脚跟后脚”。它是我从周耀旗教授撰写的一篇名为《写好英语科技论文的诀窍：主动迎合读者期望，预先回答专家可能质疑》³文章中总结而来的。我建议大家仔细看看这篇文章，每次读一遍都有新的收获。为了做到第二点，周教授在这篇文章中强调了信息的流动，并提出了两个建议：

建议一（段落）：每一段的第一句要告诉读者这一段是讲什么的，每一段的最后一句应该是这段的结论或者告诉读者下一段要讲什么。

建议二（句子）：句子是文章最小的功能单元，最容易理解的句子是整句话都在说读者容易理解的东西，但我们写文章的目的又是向读者传递一种新的观点，因此难免存在对于读者而言是从未见过的信息。为了让读者能够顺利地理解我们要表达的观点，一个好的句子应该是从读者熟悉的信息开始而以新的信息作为结束，并且要在它们之间平滑地过渡。要做到这一点，每当写一个句子的时候，都问问自己：“我以前有没有提到过这个概念”。下面是一个非常经典的例子：

“Samples for the 2-dimensional projection of kinetic trajectories are shown in Figure 7. The coil states are loosely gathered while the native states can form a black cluster with extremely high density in the 2-dimensional projection plane.”

上面这段话中，第一句话和第二句话的信息就无法流动了，因为“The coil states”不知道从何而来。如果改写成下面的句子会让人更加明白：

“Kinetic trajectories are projected onto X and Y variables in Figure 7. This figure shows two populated states. One corresponds to loosely gathered coil states while the other is the native state with a higher density.”

在这个新的段落中，新插入的第二句话可以使得每句话都能从旧信息开始而从新信息结束。第一句和第二句以“Figure”相连接，第二句和第三句以“two states”相连接，而新的信息“coil states”则出现在第三句话的最后。

从上面的例子可以看到，新旧信息的颠倒会让文章晦涩难懂。对于自己来说，所有内容都是熟悉的，不容易区分哪些是新的哪些是旧的。因此，在写每句话的时候都问一问自己，这些词汇在前面有没有被提到过，一定要把提到过的放前面，没提到过的放后面。

表 1: 常用的逻辑连接词

however	also	afterwards	furthermore	in contrast	unfortunately
in addition	consequently	moreover	unlike	similarly	alternatively
in order to	despite	for example	compared with	therefore	hence

除了注意新旧信息的呈现顺序外，一个非常有用的策略是使用连接词（常见的逻辑连接词见表 1），用好连接词能够使得句与句之间层次清楚，意思明确。下面是几种常见的连接词使用场景：

- 按照时间顺序叙述文献：最早的文献可用 Bob advocated it for the first time；接下来 Then Alice further demonstrated that ...；再接下来，Afterwards, Susan ...；如果还有的话，More recent studies by Mike ...

³https://www.journaladvancedultrasound.com/attached/file/20180802/20180802200904_688.pdf

- 叙述对立观点: Bob puts forward that ... In contrast, Alice believes that ... (or unlike Bob, Alice suggests that ...)
- 叙述相近观点: Bob suggests ... similarly (alternatively), Alice ... (or also, Alice or Alice also does ...)
- 叙述递进观点: Bob ..., furthermore (further, or moreover), Alice ...

2.2 技巧二：学会控制信息量

除了论文的强逻辑性外，我认为一篇好论文的第二个特点是能够有效地控制呈现给读者的信息量。而信息量的控制也还是从论文的命题、段落、以及句子三个方面来控制的。

首先，在命题上，一篇论文应该只提出“一”个观点，如果论文里的观点太多，不但不好写，出问题的概率也大，最重要的是读到最后，读者可能也不知道你要说什么。

其次，在段落上，每一个段落都应该只讲一个故事。在一个段落里如果呈现过多的观点反而会让读者很难知道这段想要表达什么，我应该记住什么。我在写博士论文的第一篇文章时也犯过类似的错误，当时我的导师曾一峰教授在审阅我的文章时指出了我把很多句子都糅合在一起并全部塞在一个段落，这样子不仅让读者读的非常累，并且很难看得懂你想要表达的意思。

最后，在句子上，(1) 切记使用太长的句子，尤其是加了特别多定语从句的句子。过多的长难句看似炫技，实则严重阻碍了读者阅读的流畅性，这种做法也是在疯狂地试探审稿人的耐性。(2) 力求简洁明了，用词干净准确，可以对比下面两个句子。

- (修改前) Since **there exists** rich information **that helps alleviate** data sparsity, scalability, and prediction (classification) quality ...
- (修改后) Since **exploiting** rich information **can alleviate** the problem of data sparsity, scalability, and prediction (classification) quality ...

说实话，在句子的简洁性和用词准确性上，我目前做的不够好。它是一个长期积累的过程，需要我们多阅读，多思考，多积累。

2.3 技巧三：语气要肯定，表达要清晰，意思要完整

在写作语气上，尽量不要出现助动词。肯定即肯定，否定即否定。例如：Such a function ~~may not~~ (is not sufficient to) measure ...

在词汇指向上，指代事物要清晰。我们可以通过对比下面的句子来直观感受一下：

“in which predicting student performance that helps plan educational interventions is a central topic of research.”

预测学生的成绩来帮助学生制定一些教学辅助策略是研究的主题，但这个 research 是什么呢？

这句话并没有清晰地表达出来。如果改写成下面的句子会更合适一些：

“in which predicting student performance that helps plan educational interventions is a central research topic in intelligent education systems.”

加上了“in intelligent education systems”能让读者明白原来研究的主题是智能导学系统，这样子句子表达出来的意思就更加完整了。

另外，尽量少出现如“it”、“this”、“that”等模糊指示代词。例如：It (The previous work) does not mention how to ...

在表达完整性上，特别是对方法的描述上，如果篇幅允许，尽可能详细说明。请看下面的一个例子：

“Algorithm stop criterion is set to the relative error less than ϵ ($\epsilon = 10^{-3}$), and the maximum iteration is set to be 500. We use the MATLAB *rand* command to initialize the algorithm, and the results are averaged after 20 repeated trials.”

在第一句话中，算法是如何停止的只是泛泛而谈，如果再进一步描述如何操作的会更好一些：

“Let \mathcal{R}_i is the predicted rating data at the i -th iteration, the stopping criterion of our algorithm is set as follows:

$$\frac{\|\mathcal{R}_{i+1} - \mathcal{R}_i\|_F}{\|\mathcal{R}_i\|_F} \leq 10^{-3}.$$

The maximum iteration is set to be 500. We use the MATLAB *rand* command to initialize the algorithm, and each result is the averaged value among 20 repeated trials.”

3 科研论文写作心得

心得一：不要匆匆而就地就把第一稿就给导师。很多人习惯了把写完的东西没检查就直接交给导师，寄希望于让老师给挑出所有的毛病，包括英语说不清楚、拼写错误、格式不规范等低级问题。要知道老师是非常忙的，当他看到一篇充满了各种低级错误的稿子之后，只会令他打消继续读下去的念头，然后会放下文章先去做其它事情，结果到最后还是耽误了自己。脾气好的老师会自我安慰说这是锻炼心智的过程，脾气差的老师可能就……没有然后了。令我印象非常深刻的一次是我在读硕士研究生的时候，我交给陈碧连老师的第一稿论文被批的还不如草稿。陈老师让我坐在她旁边一句话一句话地翻译给她听，有些句子甚至连我本人都不明白什么意思，当时真的是如坐针毡、汗流浹背。

因此，无论如何尽自己最大的努力多检查几遍，写的再清楚一点，尤其是格式规范问题、语法和拼写问题、图表的呈现问题（字体是否统一、大小是否合适等）。这些基本的事情是在我们手上就可以做完的，而老师的职责是从宏观上对文章进行“鸟瞰”。如果时间充裕，建议大家可以把写完的文章先放上几天，当你回过头来重新看的时候，你可能会有一种“这是什么鬼”的冲动。

心得二：不要过分地王婆卖瓜、自卖自夸。避免使用带有情感的形容词，如“obviously”、“exciting”、“remarkable”、“significant”等情感词汇来描述你的方法和结果，而要通过方法或结果的描述来使读者体会到你的文章是“exciting”的。我曾经收到过一个审稿人对我的一篇文章中的一个句子的建议，可以仔细品味一下：

I would be very careful in making a statement like this one: “The experimental results show that our WCF model **outperforms all existing similarity methods.**”

Because other scholars use clustering methods to group similar users and then perform collaborative filtering, they might achieve better results. Therefore, I would formulate the sentence in a more neutral way.

心得三：规范写作、细心修改。我在写作的时候，经常写着写着会发现某个地方不符合规范（比如名词不统一、格式不一致等），但是现在停下笔来再去修改这些东西不仅麻烦，而且会打断我目前的写作思路。偷懒的办法会告诉自己：“我先写吧，写完最后再统一修改”。但是抱有这种想法时往往导致文章写完了，这些重要的东西也忘记了，结果到最后提交时大家都没发现这些致命的错误。我的建议是先把目前的思路用注释的方式写个纲要，接着立即停下笔来去处理这些规范（如果不怕麻烦可以拿个笔记本记下这些规范要点，这也方便修改文章的时候当作 check list），然后再继续写作。因为写文章不是一天两天就能搞定的，随时保持文章的规范性也能够方便自己迅速进入写作状态。

心得四：自己的论文，自己负责到底。

4 科研论文自查顺序

无论是当你完成论文初稿准备交给同学或者导师帮忙修改，还是最终定稿准备提交论文的时候，我建议可以按照下面的顺序再自查一遍论文：

1. **Related work.** 文献是否充分，是否梳理清楚已有工作的不足。

2. **问题定义、模型.** 确保模型的细节叙述到位。

- 第一次出现英文缩写的地方一定要写出它的全称。
- 仔细查看模型内涉及的每个符号的含义是否清晰明了（千万不要含糊其辞），是否有取值（区间）且数值是否正确（有没有重复使用），注意哪些符号只是参与暂时的迭代运算，这些符号尽量不要在其它部分赋予具体的含义。例： $\|\mathbf{X} - \mathbf{E}\mathbf{U}\|_F^2 = \sum_{i,j} (\mathbf{X}_{ij} - \sum_{l=1}^L \mathbf{E}_{il}\mathbf{U}_{lj})$ ，这里的 i, j, l 都是暂时参与迭代运算的符号。
- 算法的伪代码注意检查输入、输出、初始化的参数是否正确。每个句子后面建议加分号，且叙述力求精炼，符号力求简洁。

3. **Introduction.** 多读几遍，逻辑要自洽。

4. **实验部分.**

- 图（表）的标题要明确传达出它的含义，注意指出数据是在什么样的条件下产生的（范围限定）。
- 所有的图表要用矢量图，贴到文章中，确认大小合适、清晰、对比度足够，同时也要考虑在黑白打印的情况下是否也能够看清。
- 图和表在文章中引用的位置是否是正确，不要文不对题。
- 图、表、公式在文章中引用时的缩写是否正确且统一（如 Figure 还是 Fig.）。

- 图表的内容是否表达正确（很容易出现改了一处图表之后，忘记在其他位置也做改动的情况）。
- 所有的数据是否都是正确且完整的。
- 字符、图表的间距调整，该大要大，该小要小。

致谢

首先感谢我的导师曾一锋教授对我每篇论文的敦促和指导，他是一个再忙也会腾出时间和学生们一起讨论和修改文章的学者，我从他的身上学到了许多治学精神。其次，要特别感谢陈碧连老师，她自我硕士阶段开始便手把手地教我写论文，在和她一起修改过的大大小小的论文的过程中，我也逐渐明白了如何才能写好一篇文章，应该要注意什么地方，要往哪个方向去努力。此外，还要感谢杨帆老师对我写作上的帮助，尤其记得当我的博士论文的第一篇文章被拒稿时，他和我通了数次电话，帮我复盘并和我一起修改文章，那是第一次有种茅塞顿开的感觉，内心直呼“原来是要这样写”。同时，我也要感谢我的师妹朱晋蓉对第1节提供的建议。最后，这篇指导还总结了许多前辈的经验，如果有雷同的地方且我没有标注出处，首先深感抱歉，并及时与我联系。