

Academic Paper Writing for Starters

徐悦甦/西安电子科技大学计算机学院

三个基本原则：专业、标准、统一。一个目标：经过一次论文的修改，第二次写论文的时候可以做到习惯成自然。

一、语言层次上的问题

- 1) 拼写问题，如果用 word，那么将 word 里的拼写检查选项全部打开，如果用 latex，会有默认的拼写检查，拿不准的就查字典。
- 2) 语态问题，有选择的情况下，在摘要和 Introduction 的贡献论述这两处，能用主动用主动，不要老用被动，没有必要用大量的被动，大量的被动会使用不知道作为论文主体的作者们的贡献到底在哪里？
- 3) 在摘要与叙述贡献时，大胆地用 we，不要担心用 we 显得主观，因为是谁提出的模型，是谁设计的算法，就是作者们，所以 we 在这里就是代表作者们。不要老用 this paper 怎么样，或者被动。在这里做一个强制规定：摘要/Abstract 里，禁止用被动，一句被动也不允许用。需要用被动的句子，老师来给你们改成被动，同学们一个被动也不要写。但是不需要用 we 的时候，就不要用 we。在阐述模型本身包含什么，应该怎么样时，就不要引入 we，因为这与作者们没有关系。同时，在阐述模型时，主观性的情感也不必进入，比如：The distribution $p(y_{i+1}|x, y_c)$ we are interested in is a conditional language model based on the input statement x . 这一句中，就没有必要引入 we are interested in
- 4) 在叙述模型的特点时，不要把 we 牵涉进来，一定要搞清楚是不是 we 要怎么样，一般情况下，不是 we 要怎么样。这个问题很容易犯，而且同学们不容易发现。比如，The proposed framework uses the intersections of affected links to determine the targets' position, so we can avoid the noise influence. 这里就不是 we can avoid, noise 的 influence 是这个问题或者勉强说这个框架、模型需要 avoid, we 能 avoid 什么呢？we 能避免不要把句子写错了。
- 5) 时态问题，除 related work 外，其余均用一般现在时与现在完成时即可，除了在实验部分极个别的地方可以用过去式
- 6) 大小写问题，论文中不要频繁地出现大写，会使得论文显得很乱。
- 7) 单复数问题，单复数后面的动词格式要注意保持一致。Let something do, 不是 let something does
- 8) 标点符号问题，不要使用中文标点，比如在 Latex 里，双引号是 `` ` `，不是两个引号，英文里没有书名号，也没有长破折号，也没有顿号。
- 9) 文体问题，比如句号用中文的，句号和后面的单词没有空格这些容易不注意的问题，要遵循 IEEE ACM Springer Elsevier 等出版社的格式要求。
- 10) 正式的英文科技论文，强烈不建议省略，如宾语从句的 that 等，不是因为较真，而是因为省略会使得你的主从句论点不容易分开，或者到底是要强调主句还是从句分不清。而

且有些同学第一次写论文习惯写得句子比较长，省略了引导词会进一步使得从句分不清

11) 在摘要、contribution summarization、conclusion 在叙述自己做了哪些事的时候，不要用 can be, may be, 做了什么就是什么，实事求是大胆地用 is, are。已经做出来了再用 can be, 意思就是没有做、在计划做，这是矛盾的，所以不要用 can be, may be。

12) 用 Latex 写论文的时候，不要从第一行开始一直顶格到最后一行，根据 section 的层次有个递进。全文每一行，包括公式、表格都不要一直顶格。

13) 论文文件夹命名的时候，不要只写个 1，或者写上一两字母，长一点，有意义一点，比如 The manuscript, 或者 paper for aaai。

14) 不要写 We believe, We think, 我们 be 不 believe 没有意义，关键是数据观察，理论推导或者实验结果 be 不 believe。如果可以 We believe, 那最好的应该是 We believe that our paper should be accepted.

15) cannot 在学术论文里一般是连起来写，不分开写 can not

16) 不要出现 this means that 这种句式，很难辨别在说哪个和哪个

17) 比如用 such as, 不要用 like, like 是口语表达方式。也不要 Of course 等口语词组

18) 不使用 etc 表示等等，这是口语表达方式；不使用 aka, 使用 that is, aka 是口语表达方式，同时 aka 还是缩写

19) 分清楚 such as 和 including, such as 是举例，including 是枚举/包含

20) 不要使用 namely, 用 i.e.,

21) 不要用 And 开头，And 开头是口语习惯

22) 表达与 XXXX 不同时，不要用 unlike, unlike 和 like 一样，都是口语表达方式，用 different from

23) 全文的正文字体要统一，除了个别的图片上的标注，统一为 ACM 或 IEEE 或 ACL 或 Springer 等模板中的字体，一般都是 Times New Roman。正文行文中不要出现其它字体，标题如果有需要其它字体，按照模板即可。

24) 每一行的第一个单词的首字母要大写. Abstract. 后面第一个单词的首字母要大写，Section X.X 后面的第一个单词的首字母也要大写

25) 不要把句子写得太长，同学们第一次写论文时，习惯先构思中文然后翻译成英文，然后就往往造成一句话包含好几个从句或者子句，使得句子比较长。这样的句子一方面是很不容易让人读着读着就找不到结构了，因为结构套结构就把含义套乱了，另一方面是包含的内容过多，让人组织不起来一个完整的逻辑。一般情况下，学术英文里一句话大概包含 20~30 左右单词，或者小于 20 个单词均可，甚至写若干个包含（小于）10 个或者 10 几个单词的简单句都是可以的。不要动不动就写包含 40、50，甚至更多单词的长句子，第一次写论文的同学是很难驾驭这种 40、50 个单词的长句子的，努力把一个 50 个单词的长句子拆成两个 25 个单词的正常句子。

二、 文体上的问题

- 1) 避免口语，不允许使用任何口语，例如 OK, fine, 除了举口语的例子。
- 2) 避免俚语 (slang), 不允许使用任何俚语
- 3) 避免“名言名句”，不允许使用任何名言名句
- 4) 避免语言上的缩写 (I don't, can't, it's 之类), 缩写和省略引导词 (非限制性定语从句的 that) 都是口语表达，不能出现在正式文体中。
- 5) 避免含义非常多的短语 (take on, put up, get on, make up 之类), 尽可能用一个单个词, 含义精确, 像 take advantage of 这类就只有一个含义的短语用没问题
- 6) 避免太生僻的词语, 行业专有名词除外。有道查出来的词想一想, 是不是可以通用, 不要查出来是什么就直接贴上去。
- 7) 不允许使用倒装结构, 类似 not only but also, 不允许使用
- 8) 不能使用 the more the more 句式, the more the better/more/...都是口语表达方式
- 9) 数学符号要书写规范, 例如乘号就是叉或者点, 不能用雪花
- 10) Conclusion 不是把 Abstract 再抄一遍, 是写结论, 就是我们提出了 (注意这个“了”) 什么, 得到了什么实验结果, 说明了什么, 一般是说明论文里提出的猜想、假设、推断、方法构建思路等是正确的。注意, 不是把摘要再抄一遍。
- 11) 英文期刊和会议论文的参考文献, 不要加[C][J][D], 那是中文期刊的参考文献引用格式的一部分, 英文没用[C][J][D]。再者, 先搞清楚[C][J][D]代表什么。
- 12) 限制性定语从句用 that, 非限制性定语从句用 which, 限制性定语从句不要用 which, 这样可以明确将两种区分开。写学术论文的原则不是只要语法对就可以, 和六级作文的关注点不同。用 which 来引导限制性定语从句语法上没问题, 但是会造成不必要的从句性质模糊。
- 13) 参考文献的格式要统一, 要符合会议或者期刊要求的格式, 不要一会名在前, 一会姓在前。或者一会 In Proceedings of/XXX journal, 一会又冒号: Proceedings of/XXX journal。不要一会用引号把论文题目引起来, 一会又不用引号引起来。
- 14) 是 as follows. 不是 as following, 也不是 as followed。as follows 是一个省略句, 后面用句号完结, 不是冒号。
- 15) 英语里有分号; 在作文里也使用, 但是一般情况下, 在学术论文里, 没必要使用分号。
- 16) section, figure, table 一般都用大写, 也可以用小写, 但全文要统一, 要么就都大写, 要么就都小写。不要一会 Section 2.3, 一会 section 4.1, 一会 Figure 4, 一会 figure 2。
- 17) 论文里的每一句话, 从题目到 conclusion 最后一个., 都需要自己写, 一句话不管在哪个 Section, introduction, related work, the proposed method, experiment, conclusion, 但凡有超过 2/3 以上连续的部分与已发表的论文有重复的句子, 这一句就会被认定为抄袭。超过 1/2 以上 (不一定连续) 的部分与已发表的论文中的某句话重复, 会被标出存在重复的可能性。
- 18) 是 training set, test set, 不是 testing set, 测试集是 test set。

19)极少可以使用 prove 或者 proved, 一般使用 verify, 我们研究的问题很少通过证明(prove)来论证, 只能通过实验验证(verify/testify), 所以不要使用 prove, 使用 verify/testify

20) 参考文献不能引用维基百科等百科, 要引用经过评审可溯源的会议论文、期刊、专著等, 尽可能不要引用 axiv 上的论文, 特别是 axiv 上的论文已经被会议或期刊录用了的话, 要引用录用版的。可以引用网站 (数据集来源)

21) Related work 中叙述之前学者的工作时, 用过去式。特别是, 一旦使用过去式叙述某个工作后, 后面的句子都用过去式。一般不要出现一个句子里, 有过去式, 有现在式。

22) 参考文献主要包括期刊论文和会议论文两种, 当然也可以包含书籍、网址等。对于期刊论文, 大部分都有期号与卷号, 然后后面跟着页码。对于会议论文, 一般没有期号和卷号 (有一小部分会议是由期刊出版, 可能会有, 大部分会议论文没有)。对于会议论文, 一般都是写 In Proceedings of XXXXXXXX, 因为会议上发表的论文是包含在会议论文集 (Proceedings) 里出版

23) 图的标题和表的标题, 以及表中的文字要比正文小一号

24) Latex 的.bib 文件里, 下面 4 项不用保留。

```
editor = {xxx and xxx and xxx}
timestamp = {Wed, 16 Oct 2019 14:14:55 +0200},
keywords = {XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX}
bibsource = {dblp computer science bibliography, https://dblp.org}
```

25) 使用 if XXXXX, ~~then~~ XXX 时, 不需要加 then

26) 如果使用 Word 作为论文写作的工具且使用 Word 的插入→公式, 插入公式时, 不能直接插入公式, 因为字体不是 Times New Roman, 不符合学术论文公式字体规范。故需要使用以下操作: 在公式面板中的前面找到 abc 文本, 然后点击, 然后在开始面板的字体中选择 Times New Roman (因为 Word 中的插入公式默认的字体是 Cambria Math), 然后再点击斜体直至变为斜体 (一般需要点两次)。需要注意的是, 点击斜体的时候要注意, 括号、数字上下标等不需要斜体, 故如果公式中存在括号或数字上下标等不需要斜体的符号, 需要分开点。

27) 要分清楚国际会议与国际期刊, 国际会议的名称要写全, 比较 WSDM, 不要只写一个 Web Search and Data Mining, 要写 International Conference on Web Search and Data Mining. 另外, 如果页数不够, 可以写简称, 但是要写标准的简称, 不要自己想怎么简就怎么简。比如 PAKDD, 就是 PAKDD, 有的同学写 PACKDD, 没有这种写法。ACM, IEEE, IET 这些一般都是全大写, 不要 Acm, Ieee 这样写。

28) 参考文献的作者列表要写全, 不管有多少作者, 哪怕有 20 位作者也要一一写出来, 不能省略。论文最后的每篇参考文献的作者列表里不准使用 et al.

以上三点为论文的基本写作要求。但是, 各位请注意, 一个高水平的论文, 上面是基本要求, 要出好成绩, 重要的是论文的表现力与张力, 那么表现力与张力怎么来体现呢, 参照以下要

求。

三、文章组织与表达上的问题

- 1) 图一定要清晰，大小适中，位置摆正，图里面的标注字体不要太小，否则审稿人看不清楚。
- 2) 图中的内容，包括点、线、区域，尽可能以标注、题注、小注、箭头标注等形式予以明确地说明，直接将 reviewer 的眼睛定位到图的核心内容上
- 3) 图的标题放在下面，表格的标题放在上面（不要再有问题）
- 4) 图、表、公式在给出之前要有前导说明，给出之后，一般要有后导解释。不允许出现，即没有说明，也没有解释，直接给出
- 5) 公式的书写要标准、规范，符号使用要符合数学或行业标准，避免随意使用，避免在论文中使用超大量的英文（比如 20 多个）、希腊文作符号
- 6) 解释公式后面跟着的 where 要小写，不要大写 Where，这是完整的一段内容，没有另外分段，公式也是这一段的中间内容。
- 7) 除摘要、Introduction、Related work 等部分，主干解决方案问题，不允许出现一页上全是字的情况，要么有图，要么有表，或者有公式
- 8) 图与表的样式要正确，比如什么时候用拆线图，什么时候用柱状图，学术论文中除了数据分析之外，一般不用饼状图
- 9) 所有多于一个的观点或者从图、公式、表中得出的结果，都尽可能明确、醒目地在文中以分点的形式说明，避免挤在一起
- 10) 好好使用 matlab、几何画板、visio、ppt 等工具，要让论文看得饱满、立体、丰富，不要用 word 画图。
- 11) 关于标题，标题要以名词短语、动名词短语的形式展现，里面完全可以有起修饰作用的形容词、动词，但不要以动词或形容词开头
- 12) 要强调自己的方法（the proposed method, our proposed model），要明确地将自己 propose/design/develop 的方法或模型，与参考的基准方法区分开来。但是要注意，method, algorithm, model, framework 这四种贡献形式，到底什么是 design 出来的，什么是 develop 出来的，拿不准就用 propose。还方面的词还包括 study, explore, investigate 等
- 13) 不需要大段大段地解释基准方法，更没必要用大量的标准公式去推导基准方法，适当解释即可
- 14) 要避免大量地在文中使用大写与复数，大量的大写会让文章看起来不整齐，大量的复数会让人读得很累。
- 15) 避免频繁地使用 uncertain word，像 may, maybe, perhaps, sometimes 等等，这些 uncertain words 会降低你的论点或者观察的可信性。实在要用，就用 is likely to, probably 怎么样怎么样。当时，may, maybe 也不是不可以用，但不要太频繁的使用。

16) 每一句英语句子要写对, 要有主语、谓语和宾语, 这是初中英语的问题, 但是几乎改的每篇文章都有这类问题。例如: Then select the words according to the importance and use them as the feature of the dialogue. Finally, perform the feature coding and the representation of the dialogue. 这两句都没有主语。

17) 论文正文中慎重指示代词, 例如 it, them, they, this 等, 很容易就不知道指到什么地方去了。哪怕把上一句的主语再重复一遍, 也不要用 it, them, they。

18) 半句结束之后的逗号, 整句结束的句号紧跟前面, 不是后面, 也不是夹在中间

19) 论文中的分条全部要用 1)、2)、3) 或者 1.、2.、3., 不建议用圆点或者其它符号, 那些主要是 PPT 里用的

20) 同一个概念或技术点, 在整篇文章中的表述要统一一致, 不要一会这么说, 一会那么说。例如, 相似性计算, 统一就是 similarity computation, 不要一会 similarity calculation, 一会 similar relation computation, 一会又 similarity extraction。不要一会 document summarization model, 一会又 document summarization algorithm, 一会又 document summarization strategy。再比如, 一会说 document summarization, 一会又说 document digest。这里用词的多样性完全体现不出英语水平高, 反而很明显地说明英语非常稚嫩。

21) 两个独立的句子之间如果组成一个完整的句子, 中间需要用连词, 最简单的, 用一个 and, 初中英语, 但是几乎每篇文章都存在这个问题。

22) 论文最核心的还是自己的 idea, 论文要围绕着 idea 来写, 不要写得太散, 东一点西一点, idea 是一个整体, 不要有什么贡献就罗列什么贡献, 贡献太多就写两篇论文。没有人会纯粹因为一般性的贡献的数量多就录用一篇文章, 更何况按照经验, 第一次写英文论文, 大多数同学对于自己的贡献根本表达不清楚。

23) 用 First, Second, Third, 没有必要用 Firstly, Secondly, Thirdly, 本身 First, Second 就是副词。

24) 论文中不允许出现大段空白, 这里大段空白是指 5 行以上的空白

25) 在一段话中, 不需要大量地使用 Moreover, Additionally, Furthermore 等连接词, 更不要用 What is more, 不专业, 很累赘, 可以用 Also。句子之间本身随着步骤或者逻辑的展开就存在天然的递进关系。

26) Compared to 不是 Compared with (不知道为什么, 同学们这个错误就是改不过来), the same as 不是 the same with。表达连词因为时, 使用 As 或者 because, 不是 Since, since 表让步不表示因为, 是“既然”的意思。不知道为什么同学们总是喜欢用 since 表达“因为”

27) denote 不是 donate, notation 不是 note, 也不是 notion

28) 行间公式有编号就统一有编号, 不要有的有编号, 有的没编号

29) 缩写第一次出现的时候, 要有全称, 除非特别广为人知的缩写, 而且第一次出现的时候先写全写, 然后在括号里写简写

30) 矩阵符号一般大写加粗, 一般的标量就用普通字体, 向量一般小写加粗, 集合符号一般

使用大写

31) 避免用含义特别虚、特别大 (general) 的词, information, process, analysis (analysis 可以出现在 experiment 里), 这些大词传递不出太多有价值的含义, 像 modern information system 这些是可以的。

32) 实验部分不要说 we can see, 除非特别特别明显的现象, 因为有的审稿人会说: 我就是看不出来。用 It can be seen, 或者 Figure 5 shows 之类的客观表述

33) 实验结果的解释部分, 不要光写一句话描述一下折线的走势就完了, 要再写几句话, 解释一下为什么会有这个走势, 不要光写类似“先升再降”然后就结束了。

34) 得到/产生不要只知道用 obtain, obtain 也不是很专业, 要知道使用 achieve, produce, generate, yield, form 这几个词

35) 不要老用动词不定式, to 什么什么, 要学会用现在分词 ing 形式, 特别是在摘要和 conclusion 里

36) 同学们常喜欢这么写: When using xxxx to deal with the problem of xxxx, the following steps are required. using 是人来 use, 不是 the following steps 来 use, 所以这么写主语不匹配。可以改成 When xxx is used to deal with the problem of xxx, the following step are required. 初中英语, 但是每次改都会有这个问题, 次次都有。

37) 像 Introduction, The proposed method 这些一级标题, 还有二级标题, 格式要统一, 要么就只大写一个字母, 要么就都大写, 别一会大写第一个字母, 到了下一个就又都大写。

38) 不要动不动就在正文中间大写, 一个反面例子是: The whole architecture of Attention mechanism is shown in Figure 4. 'Attention'完成没必要大写

39) 总体上建设使用 Latex 作为论文写作工具, 特别对于硕士生同学来讲, 掌握 Latex 是一项可以说的技能。

40) 在 Related work 里要想说“作者们”时, 一般计算机的论文都不是一个作者, 所以用 authors, 不是单数 author

41) 同学们在使用定冠词 the 和不定冠词 a 的时候要注意, 泛指用户, 要么用 users, 要么用 a user

42) XXX based YYY, XXX 和 based 之间要有一个连字符-, 即 XXX-based YYY

43) 像 m user n item, 这里的 m 和 n 都是数学符号, 都需要斜体, 在 latex 里就是 $\$$, 类似的还包括 user i item j dimension of latent features d

四、工具类

1) 建议使用 Latex。不同的会议、出版社, 同一个出版社的会议与期刊, 往往都会有不同的论文模板。如果用 Word, 换一个模板就要重新排一次版, 特别是涉及到参考文献部分更是费时费力。故, 建议使用 Latex 完成论文, 并且使用.bib 文件配合.bst 文件完成参考文献。同时, 硕士生掌握 Latex 也是一项技能。

- 2) Latex 经常使用的编辑工具有 TexStudio, TexLive, WinEdt, 均可。TexLive 最大, TexStudio 最小, WinEdt 更新比较慢。
- 3) 使用 Word 写论文时, 参考文献的插入使用“引用→插入尾注”, 这样可以实现参考文献编号的自动更新。
- 4) 善于使用 DBLP, 非常全面的以计算机为主的论文列表数据库, 包含各类会议和期刊。
- 5) 使用 Word 写论文时, 善于使用“引用→插入脚注”, 不要自己做, 自己做的脚注一是不会自动更新, 二是不规范。
- 6) 不要使用 qq, 126, 163 这些邮箱, 最好申请一个.edu 邮箱, 或者 Gmail, Hotmail, Foxmail
- 7) 画图的时候一定要保留原图, 不要使用 matlab 或者 visio 画完图就把原图扔了。