



微信扫一扫
关注该公众号



文 | python
编 | 小铁

前言

不知诸位在科研的起步阶段，是否曾有过如下的感受：

- 总感觉自己写的论文就是和自己读过的论文长得不太一样，也不知道为啥。
- 投稿的时候，审稿人也总是 get 不到论文的核心，只揪着论文的次要细节不放。

在本文中，笔者吐血整理大量的论文写作小细节。从格式到内容上帮你提升论文的清晰度和易读性，可以显著缓解论文写完后导师不爱看，评审看不懂等症状。

本文适合对LaTeX基本语法有所了解的同学食用。

符号篇

数学公式中的字母

- 标量符号用小写拉丁字母表示，为避免混淆字母 l 和数字 1，字母 l 可用 \ell 替代。
- 有结构的值，如句子、树、图等，用 \boldsymbol (e.g. \mathbf{x}) 。
- 向量值小写加粗。拉丁字母用 \mathbf (e.g. \mathbf{x})，希腊字母用 \boldsymbol (e.g. \mathbf{x}) 。
- 矩阵大写加粗。拉丁字母用 \mathbf (e.g. \mathbf{X})，希腊字母用 \boldsymbol (e.g. \mathbf{X}) 。
- \boldsymbol{x} 组成的集合用 \mathcal{X} ($\mathbf{x} \in \mathcal{X}$)，a 组成的集合用 \mathcal{A} ($\mathbf{a} \in \mathcal{A}$) 。
- 数域用 \mathbb{R} (\mathbb{R})，\mathbb{Z} (\mathbb{Z}) 。

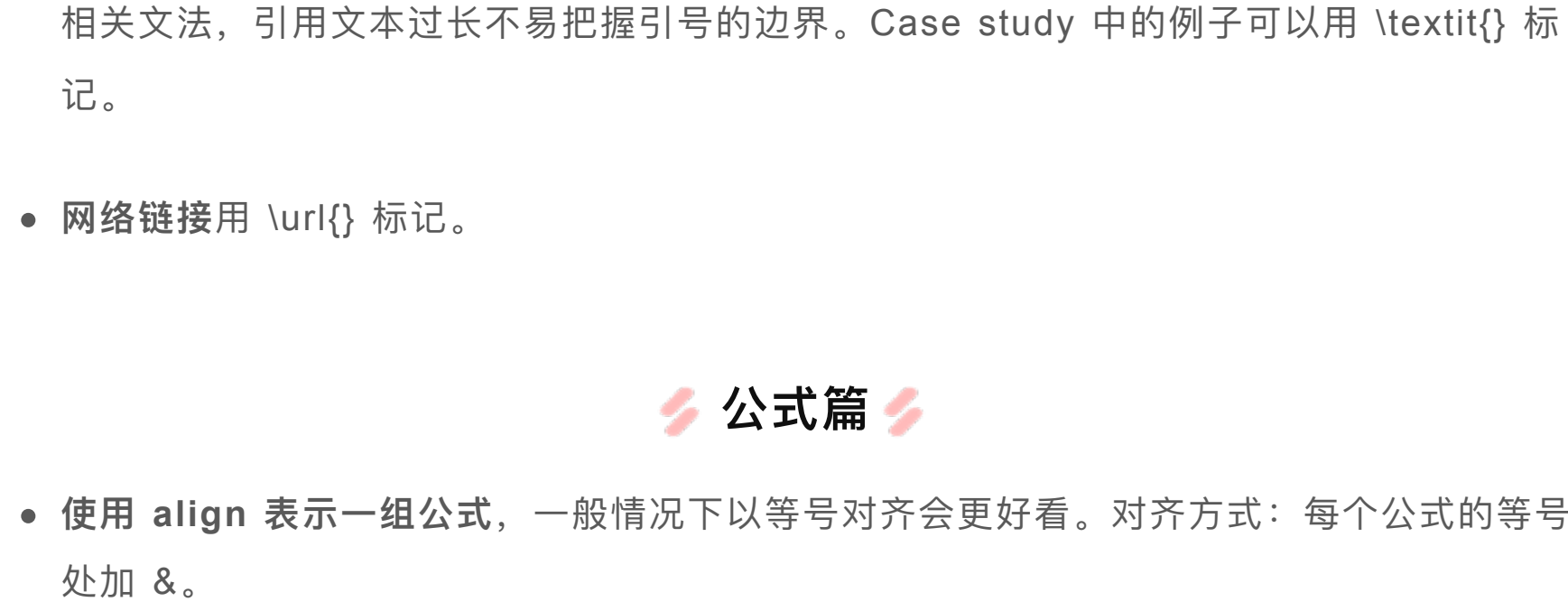
```
\begin{align}
% 标量符号:
a,b,c, \ell,k,x,y
% 有结构的值，如句子/Tree:
\boldsymbol{x}, \boldsymbol{y} \nonumber \\
%向量：注意拉丁字母与希腊字母的差别
\mathbf{a},\mathbf{b},\boldsymbol{\alpha} \nonumber \\
% 矩阵。注意拉丁字母与希腊字母的差别
\mathbf{A},\mathbf{B},\boldsymbol{\Sigma} \nonumber \\
% 集合
a \in \mathcal{A}, \boldsymbol{x} \in \mathcal{X} \nonumber \\
1 \in \mathbb{Z}, 2.3 \in \mathbb{R}
\end{align}
```

标点符号

- 数学公式中的省略号用 \ldots，例如：

$1, 2, \dots, n$

- LaTeX中英文引号的打法是：''''



- 尽量避免用引号标记较长的文本，比如 case study 中的例子。因为引号的理解是上下文相关文法，引用文本过长不易把握引号的边界。Case study 中的例子可以用 \textit{} 标记。
- 网络链接用 \url{} 标记。

公式篇

- 使用 align 表示一组公式，一般情况下以等号对齐会更好看。对齐方式：每个公式的等号处加 &。
- (建议，有争议) 只对refer的公式加编号，align中，可以用\nonumber去掉编号。
- 公式中的 softmax, proj, enc 等，超过一个字母的变量或符号，要用正文字体，即写成 \text{rm}(softmax) 或 \text{t}(FFN)。很多函数有现成的符号，例如：\arg, \max, \sin, \tanh。
- 公式中的括号，要用 \left, \right 进行标记。如 \left(\right), \left[\right], \langle \rangle, \| 这种括号也是一样的。括号中的分割可以搭配 \middle。

错误的例子：softmax 中间的迷之空格，pro和之间的迷之空格。

```
\begin{equation}
A=\mathbf{m}(\text{proj})\times \text{times } W_a \times \text{times } C_h
\end{equation}
\begin{equation}
\bar{A}=\text{softmax}(A)
\end{equation}
r_c=\bar{A} \times C_h
\end{equation}
```

部分正确的例子：还没有加粗，公式5的乘号也还没有改，只是展示一下align，以及\text{rm}的影响。

```
\begin{align}
A=\mathbf{m}(\text{text{rm}(\text{proj})})\times \text{times } W_a \times \text{times } C_h \\
\bar{A}=\text{softmax}(A)
r_c=\bar{A} \times C_h
\end{align}
```

概率相关的例子：

```
\mathbb{E}_{p(x)}[f(x)] \neq p(y|x) = p(y|x;\theta)
```

表格篇

表格的大小、换行

- 用 \centering 居中。
- 用 p{2cm} 固定列宽。
- 用 \small, \scriptsize, \footnotesize, \tiny 调整字号。
- 用 \setlength{\tabcolsep}{8pt} 调整列间距。
- 用 \multirow, \multicolumn 合并单元格。
- 用 \toprule, \bottomrule, \midrule, \cmidrule画出好看的分隔线。

示例：

		Dataset	
		Metric-1	Metric-2
Our Model	99.2	98.1	98.2
w/o p1 setting	98.1	97.1	97.2
w/o p2 setting	97.1	97.1	97.4
w/o p3 setting	97.6	97.4	97.4

表格对齐

- 展示数据的列，如果数据不等长，请右对齐。（下图左边是错误地做左对齐的示例，右边为正确的）

F-P	F-R
11.98	12.55
12.46	12.68
11.57	11.51
14.93	9.24
17.24	9.74
16.42	9.47
12.15	12.68
11.40	11.94
13.41	14.10

- 表格中的加粗请使用 \textbf{}，而非\bf。上图右边中，14.10的加粗就错误地使用了\bf，导致对齐出现问题（这一问题仅在某些模板中存在），而17.24的加粗就使用了\textbf{}。

词汇篇

避免过于绝对和模糊的表达：

以下词汇除了极特殊情况外，不要出现（→后是可供替代的词）：

- obvious (→ straightforward)
- always (→ generally)
- never (→ rare)
- avoid (→ alleviate)
- meaning, semantic, better, simple, easy, trivial ...

单复数

- phenomenons → phenomena
- 不可数名词：evidence 等。

专有名词的大小写

核心原则：符合习惯，与提出者尽量一致，第一次提时全称在前。常用词一般不大写，除非语义和词本意不同（比较少见，同时也尽量避免）。

- CNN, LSTM
- graph attention (GAT), pre-trained language model (PLM)
- FEVER, ConceptNet, SQuAD, BiDAF, FEVER score, Wikipedia

句子篇

表述尽量具体，避免贴标签

核心在于，要讲清楚提出的方法到底改善了哪里，是什么导致的这个结果。而不要总是说提出的方法提升了分数，提升了某部分分数，相对什么提升了分数。这种贴标签的话，可以有，但一定不能通篇都是。

示例1：

- 贴标签：a simple yet effective approach to solve this problem, which improves performance and robustness.
- 具体：a simple yet effective approach to help XXX to learn/recognize XXX

示例2：

- 贴标签：experiments show that our model achieves state-of-the-art performance.
- 不太具体（用于摘要）：experiments show that our model can XXX, and outperform existing methods in literature.
- 具体（用于总结）：experiments show that our model can make better and full use of XXX, especially those XXX ones, thus achieving state-of-the-art performance.

一句话尽量只说一件事

分割前：全都混在一起，读起来总有一种：“我是谁？我在哪？我在干什么？”的迷惑。

To connect XXX using relations XXX, we train a classifier on XXX, where the sizes of training/validation/test instances are A/B/C, and we finally obtain a accuracy of XXX on the test set.

分割后：做了什么，怎么做的，过程细节，结果效果。四句分开说。

We predict the relations between XXX, linking XXX. Specifically, we train a classifier using XXX. The sizes of training/validation/test instances are A/B/C, respectively. The results show that our method obtains the accuracy of XXX on the test set.

注意句子间的逻辑

主要需要明确，每句话想表达什么。表达的这个内容的前序铺垫都清楚了么？结论和条件之间是否有直接因果关系？或者比较间接但也铺垫充分了？尤其包含对于thus, therefore等因果指示词的句子，需要更加注意。

以及，段落开头的句子是否可以概括段落的核心思想？段落中每句话是否都和这个思想相关？有没有跑题？

其实，在处理段与段及章节之间的逻辑时，也都是类似的思路。

他山之石

以下是笔者多年珍藏的各路大佬给出的写作建议，这次也一并分享出来，供大家参考。

- 机器翻译学术论文写作方法和技巧：清华大学刘洋老师在CWMT-2014上做的经典报告，以机器翻译的视角，从选题开始，给大家讲解了论文写作的方法 and 技巧。

http://nlp.csai.tsinghua.edu.cn/~ly/talks/cwmt14_tut.pdf

- 如何端端正正地写科研论文？：复旦大学的邱锦鹏老师在CCL-2018的student workshop上做的报告。

<https://xpqiu.github.io/slides/20181019-PaperWriting.pdf>

- 论文写作的易读性原则：刘一佳同学，阿里巴巴达摩院的阿里星大神，在NLPCC-2018上做的报告。

<http://yjiu.net/cv/res/2018-08-19-nlpcc-sws.compressed.pdf>

- 哥伦比亚大学的Henning Schulzrinne老师的一些学术随笔，其中也有一个论文写作的collection。

<http://www.cs.columbia.edu/~hgs/etc/writing.html>

- 哈佛大学的Whitesides老师从写提纲的角度切入讲解如何撰写学术论文。

<https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.200400767>

- 如何让摘要吸引人？Nature论文摘要模板值得收藏。

<https://zhuanlan.zhihu.com/p/158574876>

- 支付宝研究员王益的建议：“学好语文，才能写好代码”（很多观点对写论文同样适用）

<https://zhuanlan.zhihu.com/p/157243326>

不管是本文谈到的细节，还是上面大佬们给出的写作建议，都只是一种风格而已，可以参考，但没有必要死守。各家有各家之言。论文写作的首要目的是支撑、凸显核心论点，而在这一前提下，保持一致最为重要。

萌新可以考虑以本文，或上面提到的某位大佬的材料中的写作风格为起点，在和导师、师兄/师姐的磨合中，逐渐形成自己的写作风格。而熟悉后也可以参考本文，吸收部分建议，融入到自己的风格中，作进一步完善。

写作最重要的还是经验。多写、多练、多总结、多思考、多听取导师的建议，相信你的写作水平会突飞猛进。让你的导师不再看你文章时心急得抓耳挠腮；让评审不会迷失在糟糕的写作中，能够一眼看出你的贡献点，打出 strong accept!



萌屋作者：python
北大毕业的NLP博士。日常写点论文，码点知乎，刷点leetcode。主要关注问答、对话、信息抽取、预训练等方向。力扣周赛第一-python选手（经常掉下来）。知乎 ID 是 Erutan Lai，leetcode/力扣 ID 是 pku_erutan，欢迎没事常来逛逛。

作品推荐

- 戳我直言，你的实验结论可能严重依赖随机数种子！
- AllenAI 发布万能问答系统 MACAW！各类题型样样精通，性能大幅超越 GPT-3！

后台回复关键词【入群】

加入卖萌屋NLP/IR/Rec与求职讨论群

后台回复关键词【顶会】

获取ACL、CIKM等各大顶会论文集！

喜欢此内容的人还喜欢

每天5分钟 | 网络自动化 (2) - JSON
网络工程师俱乐部

大高诺！论文“搜衫”，11篇不同高校论文中竟出现同一块桌布....
深度学习算法与计算机视觉

【收藏】SCI论文写作必备网站含语法绘图
地学驿站

