

如何撰写学位论文

宗成庆

中国科学院自动化研究所

开篇语

衡量一篇学位论文质量的主要标准是看其是否有重要的科学发现，是否有重要/重大的理论创新，或者是否为解决实际应用中的技术难题做出了贡献。但是，即便上述几点都具备，如果论文没有按照学术规范将研究问题、思想方法和实验结果严谨、准确、清晰地表达出来，也不能算是一篇优秀的学位论文。

提 纲

- 
1. 关于论文题目
 2. 内容布局
 3. 常见的问题
 4. 论文评价
 5. 结束语

基本要领

- ✧ 措辞简单明了，通俗易懂；
- ✧ 含意明确、聚焦，不易太大，但能概括全文；
- ✧ 尽量有特色，不易太长。

●反例：

① 基于BERT模型的问答系统研究

特色不明显

② 文字识别与文本情感分析方法研究

问题不聚焦

可修改为：面向图像文本的情感分析方法研究

③ 神经机器翻译新方法研究

题目太大

④ 面向×××的若干问题研究

题目太泛

可参考的题目

- ⑤ 大语言模型自动评价方法研究
- ⑥ 基于大语言模型的低资源语言翻译方法研究
- ⑦ 自然语言刺激下大脑的语义和句法表征机制研究
- ⑧ 面向情感支持的对话系统构建方法研究



论文题目应简明扼要地概括和反映整个论文的核心内容，**一般不宜超过25个汉字(符)，英文题目一般不应超过150个字母**，必要时可加副标题。题目中应尽量避免使用缩略词、首字母缩写词、字符、代号和公式等。

——《中国科学院大学研究生学位论文撰写规范指导意见》(2022年3月7日校长办公会议审议修订)(以下简称“**国科大学位论文撰写指导意见**”)

提 纲

1. 关于论文题目



2. 内容布局

3. 常见的问题

4. 论文评价

5. 结束语

基本要求

✧ 开门见山，尽早破题；

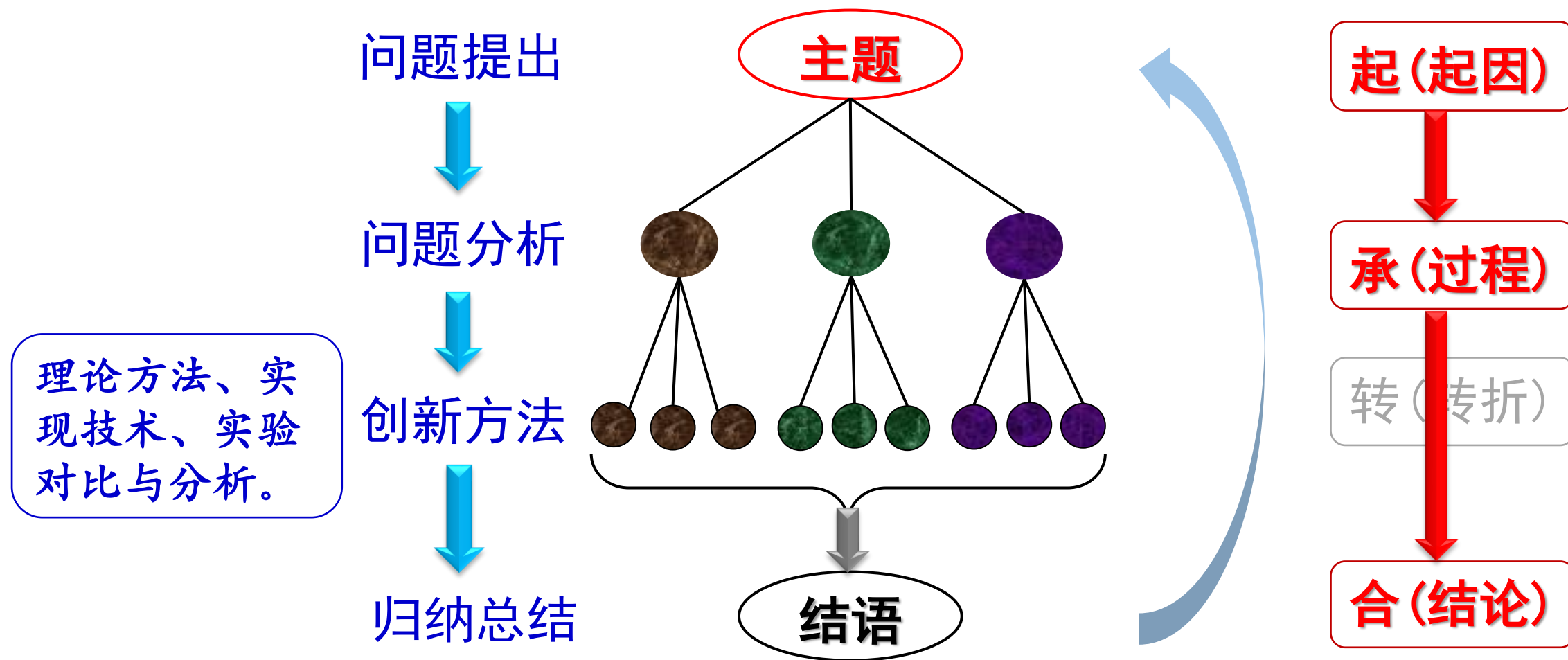
例如：“可解释性神经机器翻译方法研究”

什么叫“可解释性”？如何界定自己论文的研究范围？自己界定的研究内容和范围是否经得起考问？同行专家是否认可？

✧ 全文各章内容是有机联系的整体，有清晰的逻辑关系，尽管每一章相对独立，但都围绕同一个主题，全文构成一个完整的“故事(story)”。切忌将几个模型或者几种方法简单地拼凑、堆砌在一起。未必所有发表的“小”论文都要写进学位论文。

✧ 主题鲜明，重点突出。

阐述方式



自顶向下(Top-Down)瀑布式的展开方式

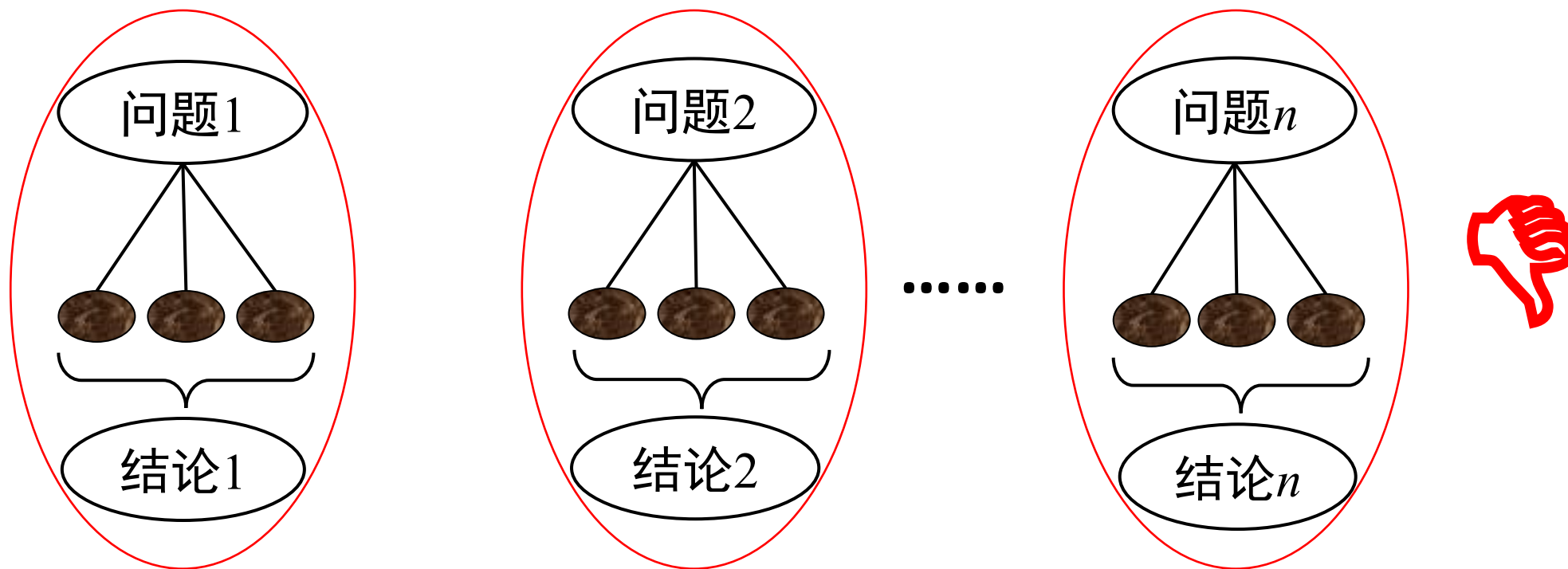
切忌...

千岛湖式的零散布局：各自独立，互不联通。



不受欢迎的表现形式-1

① 每一个问题对应一篇或几篇发表的论文，彼此之间缺乏清晰的逻辑关系，甚至毫不相关。



不受欢迎的表现形式-2

②针对同一个问题提出了两种或多种不同的研究方法，却没有阐述清楚不同方法之间的区别和联系。

- 每一种方法的优势和特点是什么？
- 对于读者而言，在什么情况下应该使用哪一种方法？
- 几种方法之间对比，结果如何？

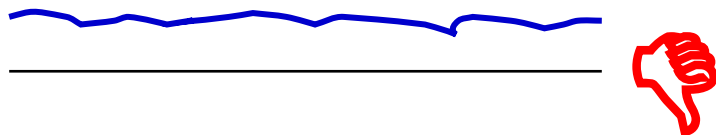
以子之矛，陷子之盾，何如？



内容分配

(1)

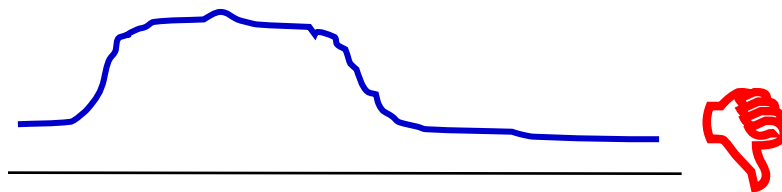
平淡无味型



各章内容平均用墨，重点不突出，索然乏味。

(2)

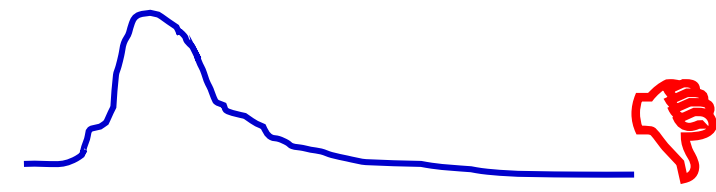
虚胖臃肿型



介绍前人的工作太多，阐述自己工作的部分过于简略。

(3)

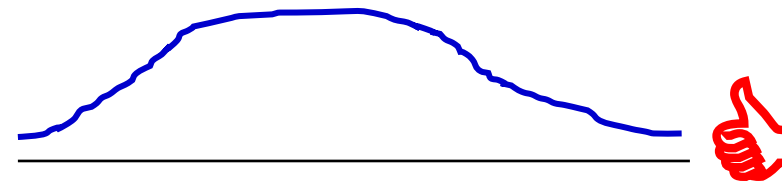
龙头蛇尾型



问题提得很大、很悬，自己的工作却没有什么实质性内容。

(4)

虎头豹尾型



对前人的工作分析深入，问题犀利，观点(方法)新颖，结语画龙点睛。

- 第1章 绪论(Introduction): 简明扼要

破题: 基本概念和主要研究内容 (做什么?)

意义: 理论意义和应用价值 (为什么?)

结构: 各章节的安排

① 博士学位论文本章一般应在6千至1万字左右(大约5 ~ 8页)。

(A4 纸, 正文用小4号字, 汉字用宋体, 英文用Times New Roman字体, 1.25倍或者1.5倍行距。全文中最小文字, 包括图表中, 不应小于6号字(8pt)。下同。)

② 硕士学位论文的篇幅可适当地减少。

各章内容安排

● 第2章 国内外相关研究(Related Work): 深入浅出

综述: 针对拟研究的问题已有哪些工作和方法?

分析: 已有方法的优缺点对比: 已经达到了什么水平? 哪些问题已经解决了? 还存在哪些主要问题和缺陷? 最好有实验或统计数据。

归纳: 方法分类, 问题归纳(预示本文将要开展的工作)。

动因: 阐述自己的大概思路, 进行问题分解。

切忌: (1)追溯的历史太久远, “自计算机诞生以来……”

(2)大量地罗列已有的工作, 面面俱到, 涵盖范围太宽泛, 没有分类, 没有归纳, 没有自己的观点;

(3)照抄别人的文献, 大量地间接引用别人的评价。

说明:

- ①在学位论文里通常在第2部分概要性地介绍本文的思路和动因，在后面针对各个创新点进行具体阐释和介绍。
- ②博士学位论文的这一部分通常在2~3万字(16~25页)左右，硕士学位论文可适当地减少。
- ③第1章和第2章内容可以合写。

各章内容安排

- 第3~5章：论文的主体，详细阐述自己的创新工作，包括理论方法、实验过程和结果，以及实验分析。
 - 阐述自己最重要的创新，几个创新点之间有密切的关联；
 - 有充分的实验对比和结果分析；
 - 建议给自己提出的方法起个响亮的名字。



重心

说明:

- ①每篇博士学位论文的创新点不宜太多，一般为3个，不超过4个。硕士学位论文适当地降低要求。
- ②博士学位论文的该部分内容应在6~8万字左右(约50~65页)，硕士学位论文的该部分内容一般在2.5~3.5万字左右(约20~30页)。

各章内容安排

- ③这几章内容应详细阐述对科学问题研究的创新成果，各章分别介绍每个子问题(根据绪论部分的分解)解决的主要思路、创新方法和与已有工作的区别，给出详细的实验对比和分析。
- 各章都有围绕论文的主题，尤其在每章的开始和总结部分要尽量呼应论文的焦点。开章部分呼应前一章方法存在的问题或者在国内外相关研究综述部分指出的拟解决问题，“本章小结”部分应分析本章方法存在的问题，预示着下一章将要开展的工作；
 - 实验对比的基线系统(baseline)应为经典模型和已有的解决同类问题的(国际)最好方法(state-of-the-art, SOTA)，而不是很久以前的方法，更不应该是名不见经传的普通水平的系统；
 - 实验结果的分析 and 归纳不可缺少。

各章内容安排

● 结束语(Conclusion): 画龙点睛

- 归纳全文的贡献和创新, 尤其要强调所提方法对于解决同类问题有没有普适性, 如何为读者提供结论性的建议, 而不是简单地重复和罗列自己做了什么, 甚至几乎与摘要内容无异;
- 指出下一步需要进一步开展的工作和未来的研究方向;
- 通常情况下, 结束语部分应在 3000~5000字左右(约3~4页)。

切忌: (1)重复论文前面的摘要;
(2)对下一步工作的展望太虚、太大。

各章内容安排

● 参考文献(References)

- ① 列出论文中参考的主要文献，对于公知的观点、公式和数字(如 e , π 等)，不必给出参考文献，但是，如果引用了别人的观点、图、表或某一段文字，必须指明出处。
- ② 博士学位论文的参考文献通常不应少于100篇，大约8~12页；硕士学位论文的参考文献不应少于50篇，约4~6页。
- ③ 通常情况下文献的排列顺序可以由论文中引用文献的先后顺序确定，或者将中文文献与英文文献分别列出，各自按作者姓名的发音顺序和文献发表的时间排列(见后面的参考文献格式介绍)。如果研究生部有统一的格式要求，应严格按照研究生部的要求执行。

各章内容安排

● 攻读学位期间发表的论文

- ①第一作者排在前面，第二、第三等作者的论文依次排在后面；
- ②期刊论文排在前面，会议论文排在后面；
- ③英文论文排在前面，中文论文排在后面；
- ④按照发表论文的重要性，尤其对学位论文的贡献，由高到底依次排列。

(并无绝对的标准，需要自己把握。)

通常情况下，一篇理工科的博士学位论文的总篇幅大约为10万至14万字(约90~120页)，硕士学位论文为6万至8万字(约50~70页)。

提 纲

1. 关于论文题目

2. 内容布局

 **3. 常见的问题**

4. 论文评价

5. 结束语

不能触碰的红线

1. 无中生有，创新点、数据等造假；
2. 大量引用他人的工作，包括图、表、公式，甚至从别人的文献中直接截取过来粘贴在自己的论文中，这涉嫌剽窃！

(按照国际惯例，如果连续7个英文单词与已有文献中的内容一样，却没有注明引用出处，就算抄袭，即使标注了引用的参考文献，其占比不能过高。)

● 内容查重

● AI 技术生成内容反查



摘要归纳不到位

- 应说明论文的主要研究目的、内容、方法和结论;
- 突出论文的创新点或新见解;
- 中文摘要长度大约2000个汉字(约1页半), 关键词3~5个;
- 摘要内容和关键词应保持中英文一致, 英文关键词应与中文关键词对应, 首字母应大写, 用英文逗号隔开;
- 不宜使用公式、图表、表格或其他插图材料, 不标注引用文献。

详见《国科大学学位论文撰写指导意见》

内容组织不合理

- 问题不聚焦，创新方法提炼不足，整个论文像是一份技术报告或系统说明书；
- 各章内容互不相干或者关系不大，逻辑关系不清晰；
- 综述国内外相关研究时像记流水账，大量地罗列，缺乏深入分析、归纳和总结，指不出已有工作存在的问题或已有方法的局限性，对近年来最新的技术方法缺乏深入了解和分析；
- 实验对比的方法或系统不是本领域公认的，或者是很早以前的，实验对比不充分；或者不是在相同的条件下(训练集、测试集、性能指标等)进行对比实验，对实验结果缺乏深入分析。

语言表达问题多

(1) 文字表达不通顺，含意不清，出现半截句：

- ① 这些比较的都是应用在检索方面，是否在分类任务中，其他的模型是否
否能发挥相应的优势还没 ...
- ② 本论文提出如何利用条件随机场解决……问题
→ 本文提出了一种利用条件随机场解决……问题的方法。
- ③ 随着计算机网络和通信技术的迅猛发展、互联网的普及与应用，人们
提出了***系统……
→ 随着计算机网络和移动通信技术的迅猛发展与普及，***技术应运而生……

(2) 口语化严重，过多地使用“我们”

- ①本文中我们提出了一种……方法，我们首先将……，然后我们……，最后我们……。我们的方法与其他系统相比……

We first design ... We then ... We find ... We align ... Our ...

- ②为了研究这个问题，我们做了下面的分析……

→为了开展该研究/为此，我们进行了如下分析……

(3) 术语使用不规范，各章节不统一

✧ 应正确使用术语，英文缩写首次出现时应给出全拼，英文词首字母用大写还是小写，全文应统一。如：

隐马尔可夫模型 (hidden Markov model, HMM)

隐马尔科夫模型 (Hidden Markov model, HMM)

n 元语法模型 (n -gram model)

N 元文法模型 (N -gram model)

token: 词元/ 子词; “道元” ☹

transformer: 转换器; 变压器 ☹

✧ 切勿杜撰、硬造术语，而无严谨、明确的定义。如：

本文建立了更加认知完备性的……表示

(4) 夸大创新点，口出狂言

尽量不要使用“提出了…理论体系”、“解决了…问题”等表述；谨慎使用“首次提出了…”，“创新性地提出了…”。

建议使用的表述：提出了一种…方法；针对…问题，提出了一种…方法；或者：提出了一个…模型。

(5) 用词华丽，不严谨

应尽量使用严谨、朴实的科技术语和普通用词，避免使用华丽辞藻，更不容夸大其词、故弄玄虚。

书写格式不规范

- 段首应空两个汉字的位置；
- 通常表题在表的上面，图题一般在图的下面，全文应统一；图中或表中应尽量使用汉语；
- 图或表的位置不恰当，或者与正文之间的间距不协调；
- 章节和各部分的标题字号大小、字体类型应全文统一；
- 公式变量是否加黑，用正体还是斜体，大写还是小写，同类变量全文应统一；
- 参考文献格式或者正文中的引用方式不规范。

参见后面的例子

写作态度不认真

◆ 错别字太多

✧ “的”与“地”不分;

✧ 英文拼写错误, 如: split ~ spit, from ~ form, window ~ widow, alternations ~ alterations

◆ 将已发表的论文中的图或表直接剪切后粘贴在学位论文中

◆ 公式错误或者书写不规范

真实例子

①

$$f(x) = \begin{cases} (x/x_{\max})^\alpha & \text{if } x < x_{\max} \\ 1 & \text{others} \end{cases}$$

$$h_l(\mathbf{X}) = (\mathbf{X} * \mathbf{W} + \mathbf{b}) \otimes \sigma(\mathbf{X} * \mathbf{V} + \mathbf{c})$$

$$R_K, R_V = \sum_{k=i}^K \text{softmax}\left(\frac{q \cdot M[n_k]}{\sqrt{d_k}}\right) \cdot M[n_k].$$

这实际上是一个点积注意 k 最近的邻居。

$$\text{softmax}(a) = \frac{\frac{e^a}{T}}{\sum_{i=1}^n \frac{e^i}{T}}.$$

$$C_k = U \sum_k V^T \mathbb{I}$$

公式书写不规范：正体与斜体混用；乘号不规范；有多余的字符。应尽量使用中文。

正体、斜体容易混用的其它几个函数：log, lg, ln, tahn, exp(.), sin, cos, argmax, Attention等。

真实例子

$$\text{WER} = \underline{100} \cdot \frac{S + D + I}{N} \underline{\%} \quad \text{WER} = \underline{100\%} \times \frac{S + D + I}{N}$$

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}^O$$

$$\text{where } \text{head}_i = \text{Attention}(\mathbf{QW}_i^Q, \mathbf{KW}_i^K, \mathbf{VW}_i^V)$$

其中, $\mathbf{W}_i^Q \in \mathbb{R}^{\underline{d_{model}} \times d_q}$, $\mathbf{W}_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $\mathbf{W}_i^V \in \mathbb{R}^{d_{model} \times d_v}$, $\mathbf{W}^O \in \mathbb{R}^{d_v \times d_{model}}$ 是

模型参数, $d_q = d_k = d_v = \underline{d_{model/h}}$ 。

$$\underline{h_{asr}} = \text{Decoder}_{\underline{asr}}(\mathbf{x}_{<t}, \mathbf{y}_{<t}, \text{Encoder}(\mathbf{s}))$$

$$p(x_t | \mathbf{x}_{<t}, \mathbf{y}_{<t}, \mathbf{s}) = \text{softmax}(\mathbf{W}_x \underline{h_{asr}} + b_x)$$

公式格式不统一；正体与斜体混用。应尽量使用中文。

其它类似的问题：概率计算中的 $P(X)$, $P(X)$, $P(X)$, $p(x)$, $p(\mathbf{x})$ 等混用。

真实例子

②表题/图题冗长

图表中一般不要使用太多的颜色，避免花里胡哨，确保最终打印出来的纸质论文清楚，非必要也不必用彩色打印。

接下来我们将会介绍在本文的实验中使用的特征。

翻译候选	双语词典	短语表	覆盖度惩罚	长度比	总得分
Bush held a talk with Sharon	4.00	2.00	0.00	0.83	6.83
Bush held a lunch with Sharon	3.00	1.00	0.00	0.83	4.83
Bush held a talk with Bush	3.00	1.00	-1.00	0.83	3.83
Bush held a talk	3.00	1.00	-2.00	0.56	2.56

表 2.1 源语言句子为“布什与沙龙举行了会谈”时，翻译候选的特征值例子。为了简单起见，在例子中假设各个特征值的权重为 1，实际上权重值会由训练得到。“Bush held a talk with Sharon”是正确翻译，得分是最高的。“Bush held a lunch with Sharon”中错误地将“会谈”翻译成了“lunch”，所以双语词典和短语表的得分比正确翻译要低。“Bush held a talk with Bush”中重复翻译了“Bush”且漏译了“Sharon”，覆盖度惩罚的得分降低。“Bush held a talk”中不仅出现了漏译的现象，句长也过短，长度比的得分会降低。总体来说，质量越高的翻译结果的总得分也越高。

表题和图题应尽量简洁，言简意赅。解释部分可写在正文中。

真实例子

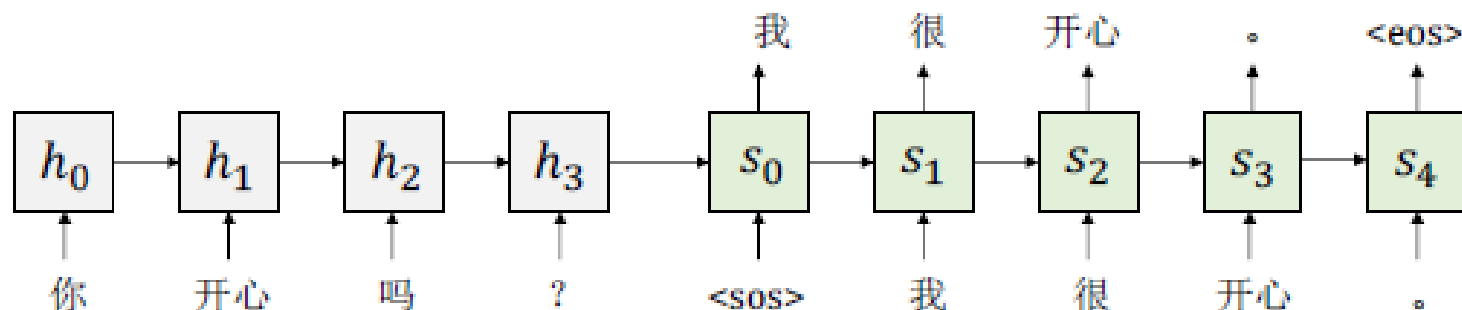


图 2.3 序列到序列生成模型结构示意图。其中 $\langle \text{sos} \rangle$ 表示句子开始的标志符号， $\langle \text{eos} \rangle$ 表示句子结束的标志符号， $h_{\{0,1,2,3\}}$ 表示编码器 RNN 的隐含层状态， $s_{\{0,1,2,3,4\}}$ 表示解码器 RNN 的隐含层状态。

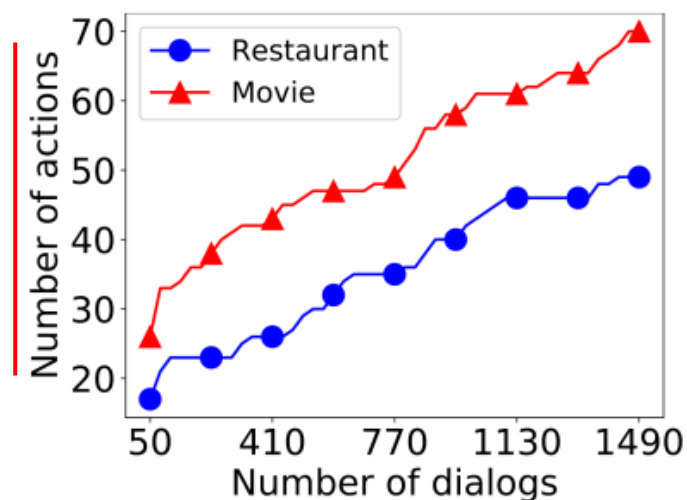


图 14 动作的数量随着收集的对话数量而增多的跟踪曲线

表 3.2 TDSLSC在数据集MR上的对比结果（主题个数为95）

Table 3.2 Performance of TDSLSC in MR dataset ($K = 95$)

Methods		Positive			Negative			All
		Pre	Rec	F1	Pre	Rec	F1	Acc
count-based	MPQA	61.40	76.90	68.24	68.79	51.40	58.70	64.15
	Liu's	73.23	62.40	67.30	67.28	77.10	71.80	69.80
	ssLDA	82.13	57.56	67.27	67.46	87.00	75.85	71.88
	TDSLSC-I	83.60	77.90	80.54	79.37	84.50	81.78	81.20
	TDSLSC-I-prior	83.38	79.57	81.38	80.51	84.07	82.22	81.82

图和表中的重要文字应尽量使用中文。



真实例子

③不同章节中公式、图和表的编号不一致，如在第3章中公式和图标记为“(3-1)、(3-2)…”；图3-1、图3-2…”而在第4章中却写为“(4.1)、(4.2)…”；图4.1、图4.2…”

正文章节题名要求最多编到第三级标题(见《国科大学学位论文撰写指导意见》)

第3章 XXXXXX

3.1 概述

… …

3.2 问题描述

……

3.2.1 YYYYYY

……

(1) VVVV

• *****

不要出现如下形式的编号：

一，XXXX

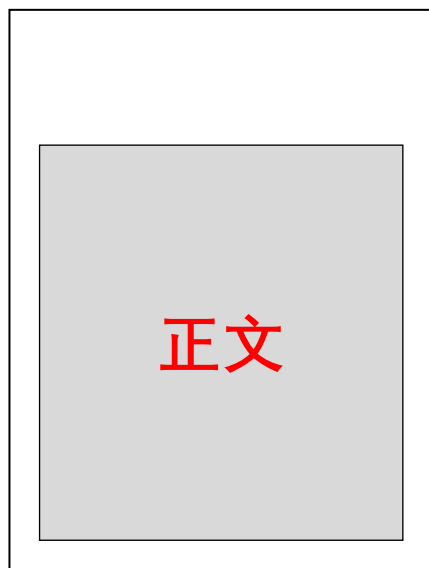
1, YYYYY

2, ZZZ

……

真实例子

④ 页面排版和字体、字号的使用不美观，例如：



.....

基本思想是，TF-IDF 分数较低的单词是无意义的，因此可以替换而不会影响句子的真实标签。通过计算整个文档中单词的 TF-IDF 得分并取最低得分来选择替换原始单词的单词。

你真棒 $\Rightarrow$ 他真棒

图 1.3 TF-IDF 单词替换

Figure 1.3 TF-IDF word replacement **【这个图太粗糙，重画！你想表达什么意思，解释一下。】**

基于简单规则的数据增强方法生成的数据多样性较差，无法适用于口语理

⑤参考文献书写不规范，例如：

OLIVETTI E, KIA S M, AVESANI P, “Meg decoding across subjects,” //Pattern Recognition in Neuroimaging, 2014 International Workshop on. IEEE, 2014, pp. 1-4

→ OLIVETTI E, KIA S M, AVESANI P. Meg decoding across subjects[C]. In *Proceedings of IEEE International Workshop on Pattern Recognition in Neuroimaging*, Tübingen, Germany, June 4-6, 2014, pp. 1-4.

XU C, BAI Y, BIAN J, et al. Rc-net: A general framework for incorporating knowledge into word representations[C] //Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. ACM, 2014: 1219-1228

→ XU C, BAI Y, BIAN J et al. Rc-net: A general framework for incorporating knowledge into word representations[C]. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM)*, Shanghai, China, November 3-7, 2014, pp. 1219-1228

真实例子

常用的两种参考文献排序方式：(1)按引用顺序；(2)按作者姓氏的字母顺序。如果是后者，建议中英文分别集中排列，参考下如下样例：

[Brown et al., 1993] Brown P F, Della Pietra S A, Della Pietra V J, and Mercer R L. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*. 19(2):263-309.

作者数大于3时，第4位及之后的作者姓名可用“et al.”形式代之。有两篇或多篇参考文献的作者姓氏相同、发表的年份也一样时，在年份后面加“a, b, c ...”等字母区分，例如：

[Abney S. 2002a] Abney S. 2002. Statistical Methods. *Encyclopedia of Cognitive Science*

[Abney S. 2002b] Abney S. 2002. Bootstrapping. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. 7-12 July 2002. Philadelphia, USA. Pages 360-367

真实例子

⑥已发表的论文列表中画蛇添足，例如：

第一作者. Application of 2019 4th International Conference on (EI Index).
最佳口头报告.

一作. “The Correlation between visual” IEEE T-PAMI, ... (AI领域知名SCI, IF=17.73, JCR一区)

已投出但尚未发表的论文，甚至正在撰写中、尚未投稿的论文，都没必要列上。

关于已发表的论文排序，请参阅本报告第20页。

提 纲

1. 关于论文题目

2. 内容布局

3. 常见的问题

 **4. 论文评价**

5. 结束语



4. 论文评价

◆ 自动化所送审博士学位论文的评阅书

学术道德评价 ↵

(一票否决) ↵

评价要素 ↵	评价意见 ↵
是否发现存在剽窃他人成果、伪造数据、由他人代写等严重作假行为 ↵	☐ 是 ↵
	☐ 否 ↵



4. 论文评价

◆ 自动化所送审博士学位论文的评阅书

▣ 评阅意见

评价要素		权重	具体得分 (百分制)
论文选题	选题的理论意义、实用价值	10%	
文献综述	反映该学科及相关领域的前人成果和前沿动态	15%	
创新成果	论文成果创新性，对学科发展、技术进步、经济建设、国家安全等方面产生的影响和贡献	40%	
基础理论和 专门知识	基础理论的宽厚度、坚实度，专门知识的系统性、深入性	10%	
科研能力	论文体现科研潜质与独立科研能力	15%	
论文写作	论文结构、撰写规范性；文字表达准确、清晰和流畅性；引文严谨、规范性	10%	
总体评价		总分	

注：“分数”栏每项均按百分制整数评分，各项满分均为 100 分。评分分为四档：大于等于 90 分为优秀；大于等于 75 分小于 90 分为良好；大于等于 60 分小于 75 分为中；小于 60 分为差。



4. 论文评价

◆ 自动化所送审博士学位论文的评阅书

对学位论文的学术评语：（对论文的学术水平、创新性做出简要评述，包括选题意义，文献资料的掌握，论文创新之处，写作规范和逻辑性等。还须明确指出论文中存在的问题和不足之处。可另附页）

↵
↵
↵
↵
↵
↵
↵
↵
↵
↵

是否同意组织学位论文答辩 ↵

评阅人签名 ↵

专业技术职务 ↵

是否博导 ↵

邮政编码 ↵

E-mail ↵

↵

↵

↵

↵

↵

↵

工作单位 ↵

评阅日期 ↵

联系地址 ↵

电话 ↵



4. 论文评价

◆清华大学送审博士学位论文的评阅书

清华大学博士学位论文学术评议书

(评议书请用黑色墨水笔书写或直接套打印, 不得粘贴)

论文题目			学生姓名	
			学 科	
评 阅 编 号		是 否 博 士 生 导 师		
专 业 技 术 职 称		评 阅 意 见 填 写 日 期		
论文的不足之处; 对论文工作的意见或建议 (请务必填写此栏)				
请在 () 中打 √, 以供参考:				
论文成果的创造性	大() 较大(√) 中() 小() 无()			
论文选题的理论意义或实用价值	大(√) 较大() 中() 小()			
论文工作中反映出的基础理论和专门知识水平	优() 良(√) 中() 差()			
论文的总结、写作水平	好() 较好(√) 中() 差()			
论文是否达到了博士学位论文要求的学术水平, 可以申请答辩? (评分时请参见下栏说明)	A(√) B() C() D()			
说明	A 达到博士学位论文要求的学术水平, 可以申请答辩; B 达到博士学位论文要求的学术水平, 但需对论文内容及文字进行适当修改后方可申请答辩; C 基本达到博士学位论文要求的学术水平, 但需对论文内容进行较大修改后方可申请答辩; D 没有达到博士学位论文要求的学术水平, 不同意申请答辩。			



4. 论文评价

◆ 北京大学送审博士学位论文的评阅书

北京大学博士学位论文评阅书

论文编号:

论文题目:

对论文的学术评语 (选题意义, 文献资料的掌握, 所用资料、实验结果和计算数据的可靠
论文创新之处, 写作规范和逻辑性等。论文中如有剽窃他人成果、编造数据等问题, 请指出。)

4-2



论文存在的主要不足之处及建议

评阅日期:

年 月 日

(可加页)



4. 论文评价

◆ 北京大学送审博士学位论文的评阅书

北京大学博士学位论文评阅评分表

论文题目		论文编号				
评 价 要 素			优	良	中	差
1、论文选题	为学科前沿，有开创性，具有较大的理论意义或实用价值。					
2、文献综述	反映该学科及相关领域的前人成果和前沿动态，归纳总结正确。					
3、创新成果	具有新的学术思路，探索了有价值的新现象、新规律，提出了新命题、新方法，创造性地解决了自然科学或工程技术或人文社会科学中的关键问题。在理论或技术、方法上有创新性。					
4、基础理论和专门知识	基础理论扎实、宽厚；专门知识系统、深入。					
5、科研能力	独立从事创造性科学研究能力。能用相关学科的思想或方法、技术解决科研工作上的难点。					
6、学术规范	论文主要内容为本人独立完成，引用他人成果有说明，论据可靠充分，逻辑严密。					
7、写作规范	论述有条理，表述清楚，文笔流畅。书写格式及图表、文字、附件、文献的引用符合规范。					
总体评价（从论文选题、创造性成果、理论水平、科研能力、学术规范等方面，给予总体评价）			是否同意组织答辩			
优（ ）			同意答辩（ ）			
良（ ）			同意答辩（ ） 修改后答辩（ ）			
中（ ）			修改后答辩（ ） 不同意答辩（ ）			
差（ ）			不同意答辩（ ）			

请您划“√”表达您的意向。



4. 论文评价

◆ 教育部抽检博士学位论文的评价表

编号：论文题目
博士学位论文抽检通讯评议评价意见表
(自然科学类)

一、分项评价

请依据评价要素，在“分项评价”栏对每个分项按“优秀”、“良好”、“一般”、“较差”四档进行评价。

评价指标	评价要素	分项评价
选题	1. 选题的前沿性和开创性； 2. 研究的理论意义、现实意义； 3. 对国内外该选题及相关领域发展现状的归纳、总结情况。	良好
创新性 及 论文价值	1. 对有价值现象的探索、新规律的发现、新命题新方法的提出等新的科学发现； 2. 对解决自然科学或工程技术中重要问题的作用； 3. 论文及成果对科技发展和社会进步的影响和贡献。	较差
基础知识 及 科研能力	1. 论文体现的学科理论基础坚实宽广程度和专门知识系统深入程度； 2. 论文研究方法的科学性，引证资料的翔实性； 3. 论文所体现的作者独立从事科学研究的能力。	较差
论文规范性	1. 引文的规范性，学风的严谨性； 2. 论文结构的逻辑性； 3. 文字表述的准确性、流畅性。	一般



4. 论文评价

◆ 教育部抽检博士学位论文的评价表

二、总体评价

总体评价分为“合格”、“不合格”两档。如果评为“不合格”则须在第三项“综合评价意见”中填写主要理由。

☐合格 ☒不合格

三、综合评价意见

请在下表中填写综合评价意见。对总评“不合格”的此处必填，给出不合格主要理由。

--



提醒：谨慎使用AI工具

论文写作过程中可以借助大语言模型等AI工具帮助润色和校对文字，但不能让AI工具成为主笔。对于AI工具处理完成之后的文稿作者必须认真、严肃地审核检查。无论使用什么辅助工具，作者对文稿内容的真实性、准确性和规范性都负有不可推卸的责任。

宗成庆出版过哪些专著？

宗成庆教授是中国自然语言处理领域的知名学者，曾出版多部专著，主要包括：

1. 《统计自然语言处理》（第1版、第2版）

该书系统介绍统计自然语言处理的基本理论和方法，是国内自然语言处理领域的经典教材。

2. 《文本数据挖掘》

聚焦于文本挖掘的技术与应用，涵盖信息抽取、情感分析、话题模型等内容。

3. 《自然语言处理导论》

面向初学者的入门教材，兼顾基础理论与实用技术。

4. 《机器翻译》

深入探讨机器翻译的原理、方法和前沿进展，涵盖统计机器翻译和神经机器翻译。

此外，他还参与编写了多部中外文专著及会议论文集，并在国际期刊和顶级会议（如ACL、EMNLP）上发表了大量研究成果。如需获取更完整的列表，建议查阅其学术主页或相关数据库。

《机器翻译》是哪一年、哪个出版社出版的？

宗成庆教授撰写的《机器翻译》一书于 **2014 年** 由 **高等教育出版社** 出版。

该书是“自然语言处理丛书”中的一部，系统介绍了机器翻译的基本原理、统计方法以及后来的神经机器翻译等内容，适合作为相关领域的学习和研究参考。

胡说八道！

提 纲

1. 关于论文题目
2. 内容布局
3. 常见的问题
4. 论文评价

 5. 结束语

5. 结束语

- 我们一向鼓励创新，但创新应体现在学术思想和技术方法上，而不是在写作规范和形式上；
- 不同学科领域和不同研究方向的论文写作风格和格式要求可能有所差异，应遵循本学科领域认同的写作规范。报告中的建议和要求只代表个人观点，仅供参考；
- 论文的选题和采用的技术路线及实现方法和要求，都应在导师的指导下完成。本报告只是从写作的角度阐述论文撰写的基本要求和应避免的问题；
- 报告中所列举的示例在很多学位论文中普遍存在，若与某篇论文中出现的问题雷同，纯属巧合。示例只用于说明问题，并无任何不良用意；
- 善于思考、观察和学习，认真对待，用心写作，是完成高质量学位论文写作的基本要求，也是必然要求。

学位论文，尤其是博士学位论文，对于绝大多数同学来说是一生中最后的一份答卷，它将伴随作者一生存放在图书馆和互联网上，甚至比一生更久远，任何人都可以对它评头论足。从论文提交的那一时刻起，作者、导师和培养单位就被牢固地绑定在一起，一荣俱荣，一毁俱毁。为人负责，为己负责，养成扎实、严谨的学术作风，必将受益终生。



谢谢！