

# 实验室研究生科研指南

宋巍

首都师范大学

wsong@cnu.edu.cn

摘抄：

- 取法于上，仅得为中，取法于中，故为其下。
- 做研究很难。但是，如果你想达到很高的标准，生活中的大多数事情都很难。
- 最高的追求是为人类做贡献。
- 真正有意义的研究如光束划破夜空，照亮人类位置的黑暗；真正有意义的科研人生是成为这束光。
- 仰望科学史上的群星，我们要努力积蓄能量，争取做一只萤火虫。
- 最高的奖励是内心的自我肯定。
- 为人类做贡献、为国家做贡献、为行业做贡献、为自己做贡献……

## 1 时间投入

做任何事情如果想达到很高的标准都需要投入大量的时间。

作家格拉德威尔在《异类》一书中指出一万小时定律：“人们眼中的天才之所以卓越非凡，并非天资超人一等，而是付出了持续不断的努力。1万小时的锤炼是任何人从平凡变成世界级大师的必要条件。”

西湖大学校长施一公也曾提到：“所有成功的科学家一定具有的共同点，就是他们必须付出大量的时间和心血。这是一条真理。实际上，无论社会上哪一种职业，要想成为本行业中的佼佼者，都必须付出比常人多的时间。……有时，个别优秀科学家在回答学生或媒体的问题时，轻描淡写地说自己的成功凭借的是运气，不是苦干。这种回答其实不够客观、也有些不负责任，因为他们有意忽略了自己在时间上的大量付出，而只是强调成功过程中的一个偶然因素，这样说的效果常常对年轻学生造成很大的误导。”

著名生物学家、美国科学院院士蒲慕明十多年前在加州伯克利大学任教时给博士生的信也非常著名。他在实验室里严格执行一系列规定，这里列举两条：

1. 每周必须在实验室里工作至少50小时，也就是每天8小时以上，每周6天。这已经大大低于我现在和我职业生涯里大多数时间里的工作强度。你们可能比我聪明，或者不想取得成功，所以我不会要求你们达到我的工作时间。等你们开始写论文的时候，可以多花点时间在阅读、写作和文献搜索上。
2. 工作的意思，就是坐在椅子上工作。不包括上网和收发与工作无关的非科研性邮件。你可以下班后呆在实验室或者回家做这些事，或者尽情去聊和科研没有关系的事。除了特殊的日子，午餐时间不能过长。我建议每个人每天花至少六小时踏踏实实地专心工作，再花2个小时以上阅读和做其他与研究相关的事。阅读论文和其他书籍要放在下班后做。

以上的例子充分说明，如果想取得学习和工作上的成功，足够的时间投入和高效的时间安排是必不可少的。当然，你可以说：“我既不是、也不想或许也不需要达到伯克利博士生的水平，更不奢望达到施一公、蒲慕明那样的高度，我为什么要这样努力工作？”没错，我希望但

不奢望你们能够达到如此高度，但是我希望你们离开实验室的时候能够比进入实验室的时候提高很多，能够带着丰硕的成果自信地迎接新的挑战。即使这样一个小目标，也需要付出时间和努力。每一个人都应该充分挖掘自己的潜力，高标准地要求自己，做到最好的自己。未来才有可能获得更多的机会，得到过得去的成绩和机会，能够为家庭和社会做出贡献。

如果你对研究实在没有兴趣，我建议你应该立即离开实验室，去尝试自己真正感兴趣的事情，不要在这里浪费时间。

在时间投入方面，要求实验室研究生至少做到以下要求：

1. 除非有特殊情况，工作日必须在实验室工作。根据我的体验，在家或者在寝室很难做到全身心的投入。更重要的是，实验室提供了相互学习、分享、讨论的学术环境以及同伴之间交流、沟通、建立友谊的生活氛围。你会发现，饭桌上的吐槽倾诉，操场上的漫步聊天是很好的减压方式和生活调剂。
2. 每天应该在8:30之前进入实验室开始工作，一天之计在于晨。每天工作净时间不应该低于6个小时（即使8:30到实验室并立即全身心投入，上午也只有3个小时左右的工作时间）。尽可能集中大块时间踏踏实实进行学习，减少碎片化时间。
3. 工作时间不允许进行追剧、看比赛、打游戏等与研究无关的工作。
4. 每周一晚上9:00之前提交周报，详细记录本周的工作内容、上周计划的完成情况、实验设置和结果、收获和问题等，以及下周计划。邮件抄送所有人。每周四进行小组讨论和论文阅读，不定时地进行一对一讨论。
5. 每三个小时检查一下微信、每天至少检查一次邮箱，如果我给你们留言，必须尽快回复，不能超过三个小时没有回音，否则沟通成本太高。除特殊情况，我尽量不在上午和你们讨论，把最好的时间用于研究。

以上只是基本要求。遇到论文截稿的日子，可能需要高强度工作。时间管理是很大一门学问，每个人都应该学习并找到适合自己的方法。每个人都要努力做到自律，养成好习惯，严格要求自己。

## 2 完成一项研究工作的主要过程

硕士论文应该至少有2到3个创新点，即要有2到3项完整的研究工作。完成一项研究工作的主要过程一般包括以下阶段：

1. **确定研究主题** 无论发表一篇高水平会议或期刊论文，还是毕业论文，都需要有一个明确的主题。一般先确定一个大的方向，而后在其中选择一个或几个具体的点作为研究的突破口。与做工程项目不同，偏向学术的科研要求研究工作具有创新性。创新性可以用两个词来概括：new + better，既要新、又要好。
2. **文献调研** 既然要“创新”就必须了解已有的工作已经达到什么水平。文献调研就需要全面地了解和研究主题相关的工作，包括数据集、方法、当前的最高水平和存在的问题与挑战等。文献调研的产物应该是一篇研究综述。
3. **实现Baseline系统** 调研结束后，应该确定当前效果最好或最有代表性的方法作为基准系统（Baseline）。Baseline是接下来研究的标杆，是判断和证明创新的参照物。新提出的方法是否创新，主要和这些Baseline进行比较，包括方法和效果。应该亲手编码实现这些Baseline，以深入了解已有方法的精髓和存在的问题。
4. **提出新主意（new idea）** 对Baseline或者自己已有系统进行深入分析，发现关键问题，提出新的研究任务或针对已有研究任务更好的解决方案。
5. **实验验证** 准备数据、编码实现新的解决方案并与Baseline比较。如果新方案取得了更好的效果，应该进一步仔细分析方案有效的原因，撰写论文。如果效果不佳，回到第4步。

6. **写论文** 研究工作最后的呈现方式是学术论文。学术论文写作是有一定规范的。需要注意的是，并不一定在所有工作完成后才进行写作。写作可以贯穿在研究工作的整个过程中。例如，文献调研的同时就可以写相关工作（Related work），提出新方案的时候就可以写（Introduction）和方法（Method）。这些初期的写作可以是草稿，不需要很规范，但对理清思路和提高完整版本的写作效率是非常重要的。草稿中存在的问题可以在整体合稿的过程中不断修订。
7. **投稿** 将完成的论文投往领域内的重要或顶级学术会议或期刊。会议审稿期通常在3-4个月，反馈较快，但是没有修改的机会。一旦审稿人认为论文达不到会议要求的水平就会被拒绝。顶级会议的录用率一般在30%以下，个别顶级会议的录用率只有10%-15%。期刊审稿期在半年以上，但是可能会给修改的机会。如果第一轮的审稿意见不是拒绝，就可以根据审稿人的意见进一步修改，并进行更多轮次的审稿。主流会议一般每年只举办一次，有明确的截止时间，而期刊可以随时投稿。如果论文被拒绝，就需要继续修改后投往其他的会议或期刊。因此，投稿并不是一蹴而就，可能需要多次投稿、修改的过程才会被接收。因此，一方面要加快研究工作进程，另一方面要珍惜每一次投稿机会。

### 3 确定研究主题

确定研究主题实际上是非常困难的一件事。初级研究生可以和导师讨论，先从一个具体的问题入手，把研究的整个流程完整地走一遍。在解决这个问题的同时，进一步拓宽和加深对相关领域研究状况的理解，找到自己兴趣点。同时看看牛人们都在做什么问题、他们怎么创新和写作，领悟其中的套路和精髓，逐渐能够做到自己提出和实现新的想法。

另外，要了解自己的优势和劣势。例如，如果数学好，可以做偏模型偏理论的课题；如果觉得数学一般，可以去深入了解具体应用问题的性质和特点，采用A+B的策略进行创新。例如，现有方法为A，但是没有考虑B问题/资源/信息，于是可以引入B来改进方法A。如果编程好，但是不善于找idea，可以与他人合作。

### 4 文献调研

确定研究主题后，便开始调研相关文献来充分了解已有工作。

#### 4.1 获取相关论文

可以通过以下方式获取相关文献：

1. **关键词+学术搜索** 强烈推荐使用谷歌学术 (<https://scholar.google.com/>) 去找论文。如果谷歌学术不可用，可以使用Semantic Scholar (<https://www.semanticscholar.org/>)。注意：不要使用百度学术等其他学术搜索。  
使用Latex写作需要论文的bibtex信息来建立文献索引。谷歌学术本来提供该信息，但经常无法访问。Semantic Scholar提供的bibtex信息经常不全，可根据论文实际信息手动补充。
2. **ACL Anthology** 自然语言处理（Natural Language Processing, NLP）领域可以直接访问ACL Anthology (<https://www.aclweb.org/anthology/>)。NLP领域顶级和重要会议（ACL, EMNLP, NAACL, COLING等）、期刊（Computational Linguistics, TACL）的每一年每一期都可以在这里找到。ACL Anthology也提供搜索接口，也可以用关键词搜索。ACL Anthology中的论文都提供bibtex信息。因此，NLP领域论文优先从ACL Anthology获得bibtex信息。如果找不到，再到Semantic Scholar去查询和补充完整。
3. **研究综述** 一个快速、全面了解某个主题研究内容的方法是读研究综述（Survey或Review）。综述一般由资深的研究人员对全面概述某个主题的研究进展，通常包括对过去一段时间内重要论文的归类、分析和对未来工作的展望。  
综述类文章可以从中文期刊找。国内期刊主要参考《计算机学报》、《软件学报》、《计算机研究与发展》、《电子学报》和《中文信息学报》等。其他的中文文章没有必要看或者简略看一下即可。中文期刊文章有新想法的少，而且实验普遍不充分。

4. **顶级会议论文集** 顶级会议每年最新的论文集集中的每一篇都应该至少快速泛读一遍，以掌握最新的前沿技术。除了ACL Anthology涵盖的NLP领域顶级会议论文，其他文本处理相关的顶级会议包括：机器学习领域的ICML、NIPS、ICLR，人工智能领域的AAAI、IJCAI，数据管理、挖掘与搜索领域的KDD、SIGIR、CIKM等。现今很多研究涉及到多模态，计算机视觉领域（Computational Vision）也很活跃，可以关注CV领域的顶级会议ICCV、CVPR、ECCV等。
5. **期刊或Arxiv** 有很多期刊也非常优秀。期刊论文要比会议论文更长、内容更丰富、详实。自然语言处理中比较著名的期刊包括《Computational Linguistics》、《Transaction of ACL》（这两个期刊可以在ACL Anthology中获取）以及《IEEE Transactions on Audio, Speech, and Language Processing》等。高水平的会议和期刊可以参考最新版《中国计算机学会推荐国际学术会议和期刊目录》。

Arxiv是活跃的论文分享网站。很多没有经过同行评审的文章或者还没有正式发表的文章会预先挂在Arxiv上。但Arxiv上的文章良莠不齐。很多人通过Arxiv来“占坑”，宣誓对某些idea的主权。
6. **参考文献** 一旦已经获得一些参考文献，可以通过阅读这些文献的相关工作部分以及它们的参考文献来获得更多相关论文。
7. **社交媒体** 可以关注知乎、活跃的领域相关公众号或者活跃的学者社交媒体账号，他们经常推荐热门的论文，可以帮助筛选和推荐近期热点，但不要盲目相信其中的吹嘘部分。
8. **问导师** 请导师推荐相关论文，或请导师帮助判断你要读的论文是否相关。一定不要羞于向老师求助，应该多问问题、多讨论，要去推动（push）老师花更多时间在你的身上，而不是等着老师push你。事务繁忙或指导很多学生的老师根本没有足够时间去指导每一个学生。谁push老师，谁就可能收获更多。

## 4.2 如何读论文

沈向洋博士做过一场线上公开课《You are how you read》，分享了他对于科研论文阅读、撰写的经验，大家可以观看参考。读论文涉及到以下一些关键点。

### ● 读什么？读论文的时候应该带着问题去读。

- **中心思想**：这篇文章主要讲了一件什么事？
- **内在逻辑**：作者怎么讲这个故事？如何引入问题、如何解决问题、如何评价效果？
- **实验手段**：作者提出的观点能够站得住脚吗？他是怎么证明的？

### ● 怎么读？不同类型的论文采用不同方式。

- **快速地读** 对于大多数论文，尤其是和自己研究任务明显不同的文章，只要快速泛读即可。读一读Title、Abstract和Introduction、浏览论文的图表结果，大致地了解论文研究的主要任务、大致方案，评价数据和效果。如果提出了你很感兴趣的新任务，可以看看数据是否能够公开获取。如果提出了一个新方法，想一想这个思路能不能迁移到眼前处理的任務上。快速地读不需要深入了解论文细节，主要目的是在头脑中留下印记，以后需要的时候可以回想起来，再来查看。
- **批判地读** 读论文要批判地读，不断地去质疑它。不要迷信顶会顶刊的论文，不要认为已发表的论文全部是正确的。应该假设自己是审稿人（reviewer），去看一看其中是否有不合理的设置或推断，或者去学习其中写得好的地方。

审稿人如何去审视一篇文章呢？顶级会议论文审稿主要考虑论文的以下性质：

- \* **创新性（Novelty）**：提出的问题或者方法是新颖的吗？
- \* **显著性（Significance）**：论文的贡献是显著的吗？
- \* **可行性（Soundness）**：提出的方法在技术上可行吗？
- \* **评价（Evaluation）**：提出的论断（claim）能够被理论分析或实验结果支持吗？
- \* **清晰性（Clarity）**：文章撰写是否结构完整、逻辑清晰、表达清楚？

在最近两年，审稿人被要求提供更多地证据来支持其对所评审论文的判断。因此，审稿人还需要提供具体意见。一般包括：

- \* 论文的优点有哪些？
- \* 论文的缺点有哪些？
- \* 论文可以从哪些方面去提高？

读论文的时候，应该从以上方面去审视、判断。注意，批判不是批评，没有论文是完美的。应该去伪存真、客观地去审视所读的论文，赞赏并学习其优点，从论文的缺点中吸取教训并从中寻找新的研究方向。

- **深入地读** 当遇到和所做问题密切相关的文章，当你已经能够判断出这些文章是优秀、可靠的，就应该深入地去理解论文的每一个细节，包括模型的理论和实现、实验数据、实验设置、实验设计和结果。这些文章将是未来工作的基础以及必须击败的对手。知己知彼方能百战不殆。

可以通过目录或一些软件帮助管理读过的以及要读的论文。论文的bibtex信息都要找到并存在Bib文件中。这些平时的小事在写论文的时候会达到事半功倍的效果。

### 4.3 写研究综述

文献调研的成果是写出一篇研究综述，对和研究问题相关的重要论文进行全面收集，对研究问题的定义、发展历史、方法演化、数据集、评价方法进行全面、有逻辑地概述。这一点非常重要。很多人做了很久，还对领域的基本常识、理论和方法不了解，这样的研究基础是极不牢固的，更无从谈及创新。如果包括为解决问题引入的机器学习方法的文章，每一个研究问题至少要涉及50篇到100篇重要论文，甚至更多。应该把这些论文的按照方法、时间、创新点等维度分门别类地整理，并用自己的话概括出来。至少要精读其中20-30篇论文，对最相关的10篇-20篇论文要如数家珍。

研究综述也可以投稿到期刊，好的研究综述有非常大的影响力。即使不投稿，这也将是未来毕业论文的第一章，更是创新的起点。每一个人都应该写一篇相应研究问题的综述文章。

## 5 实现Baseline系统

当已经了解相关工作，就应该筛选出有代表性的典型方法作为Baseline。目标就是要击败这些Baseline。一类Baseline是引用数较高的经典方法；一类Baseline是新近发表的，在一些公开数据集上取得了目前最好结果的方法（State Of The Art, SOTA）。应该动手实现这些方法并争取复现论文中的指标。

如果研究任务比较新或新提出一个新任务，直接的相关工作较少，那么应该选择更一般的相关工作中的典型方法作为Baseline。例如，提出了对诗歌表达方式识别问题，以往没有类似工作。但是，这个任务本质上可以视为文本分类或者序列标注。因此，文本分类和序列标注相关的大家普遍接受的基准模型就应该用来做Baseline。这种情况往往要多尝试不同的Baseline。

### 5.1 准备实验数据集

准备实验数据集是研究中的重要一环。

使用公开数据集最容易与已有工作比较。可以直接使用已发表论文上在相应数据集上的结果。如果有模型比较复杂且不是很重要，但又不得不引用，可以不实现该模型而直接与其在公开数据集上的结果进行比较。这种方式的优点是省事，缺点是有些文章在实验的时候采用了很多技巧导致实验结果虚高，后来者如果没有掌握这些技巧很难超越。

如果自己构建了新的数据集，应该编码实现Baseline模型并将Baseline在新数据集上的结果作为对比的基准。

### 5.2 关注开源社区

可以到Github上搜索论文题目或模型名字，查看是否已经有对应的开源项目。如果有，可下载下来作为Baseline。一般应该选择Star比较多的、大家普遍认可的项目。否则，应该自己确保采用的开源项目实现得没有问题，而不应该盲目地即插即用。

## 5.3 动手实现

当没有开源代码可用时，需要自己动手实现这些Baseline。

编程能力必须过关。很多情况下，很好的想法由于实现过程中出现失误而没有达到应有效果，与好成果擦肩而过。这是非常可惜的。正确地用代码实现想法是一切的基础。

提高编码能力最有效的方法就是多动手实践、多读源代码。现在很多开源项目的代码质量非常高，注释清楚、示例丰富。读代码不仅能够提高编程素养，而且能够真正理解一个模型，有助于提出和实现新的模型。

此外，要学会通过搜索引擎和社区论坛解决实现过程中遇到的问题，像Stack Overflow等论坛能够解决绝大多数的技术问题。

实现Baseline最关键的是要尽可能地完全复现Baseline的实验设置，包括：数据集的划分、评价指标、超参数设置、模型的实现细节等。这样得到的实验结果才有可比性。自己在记录实验结果时应该同时记录当时的实验设置细节。

如果自己实现的Baseline无法达到论文中的指标，可以发邮件向作者求助。另一个方法是将自己实现的模型应用到一个旧的、相对简单的任务上，检查其能否正常工作；或者，找小伙伴多人独立实现，如果实验结果不一致，必定实现过程中有问题。

## 6 提出新主意

提出new idea是整个研究过程中最困难也是最有吸引力的地方，一旦能够想出一个巧妙、新颖的idea，那种兴奋的感觉是非常美妙的（，虽然要做好接受导师一盆凉水的准备）。

### 6.1 创新的类型

1. **开创性的** 最大的创新是提出了过去没有的新的问题或任务，引领了新的研究方向。所谓大牛“挖坑”，大家“跳坑”。要做到这一点是可遇不可求的，需要天时、地利、人和。例如，神经网络在上个世纪已经被研究了20年，但是并没有火。伴随着可用的数据规模越来越大，计算机算力越来越强，基于神经网络的深度学习才迎来了爆发。
2. **颠覆式的** 研究已有问题，但是提出了全新的方法并取得了碾压式的性能提高，如BERT。
3. **增量式的** 研究已有问题，在已有方法的基础上修修补补，不断完善和提高已有方法。大多数的创新都是增量式的（incremental）。增量式创新也是有意义的，量变促成质变，每一次颠覆都是大量的增量式创新积累而成的。
4. **实验比较式** 还有一类研究不算完全创新，但也很有意义。有时一些新的任务刚刚出现时，很多论文可能采用不同的方法、在不同的数据集或者不同的实验设置下产生了不可比较的实验结果。此时，通过一定策略，使得不同方法能够被公平的比较，从而回答哪个方法在哪些条件下表现更好这一类的问题，可为这一方向未来的发展提供参考。

### 6.2 如何能够创新

这是一个很难的问题。首先，很多自以为的创新可能早就有人做过，只是自己不知道。其次，有了新的想法，但可能是错误的或者不可解释、说不通。最后，想法很好，但实验效果却不好。因此，创新的尝试经常是失败的。

要想成功地创新，至少应该做到以下几点：

1. **充分、深入地了解相关任务和领域** 这就是为什么我们要写研究综述的原因。此外，多想人是怎么解决这些问题的。虽然现在神经网络模型已经是主流，但强烈建议先尝试实现一个基于人工构建特征的模型，能够更加深入了解任务的特点，也有助于设计出动机明确、可解释的高级模型。
2. **交叉创新** 用A领域的方法C，解决B领域的问题D，成为第一个将C用于解决D的人。例如：文本摘要中的经典方法TextRank就是将搜索引擎采用的PageRank算法迁移过来。
3. **技术创新** 将最新出现的技术应用到目标任务。例如，最近两年非常火的图神经网络模型在自然语言处理中的应用。这需要平时关注机器学习等领域的最新进展。

4. **模仿创新** “熟读唐诗三百首，不会吟诗也会吟。”多读论文，总结创新套路。
5. **深入分析实验现象** 多分析实验结果、多标注数据，注意归纳错误类型，透过现象看本质，有针对性地思考解决方案。
5. **打破思维定势** 越是约定俗成的东西，越要想为什么要这样？这真的是对的吗？研究就是要质疑一切。
6. **多和别人讨论** 不要羞于说出自己的想法，多和别人讨论，尤其是和研究方向不一样的人讨论，往往能够碰撞出新的想法。

以上是一些必要条件，而非充分条件。要坚韧不拔、不怕失败、反复试验，思考和解决问题的过程是痛并快乐的，解决问题之后能够获得慢慢的成就感。

判断自己的工作是否有创新性的一个简单的方法是：假设你要写一篇文章介绍你的方法，给你的文章起一个名字。如果你不能起出一个名字，说明方法的创新性要么太小（trivial），要么思路不清晰（not clear）、要么不系统(not principle)。

## 7 如何做实验

能产生新想法已经成功了第一步，还要用实验去验证。实验是低年级研究生最易忽视的地方，因为本科毕业论文或项目往往对实验并不重视。实际上，实验设计和结果对论文的结论能否站住脚、论文能否被高水平会议和期刊接收都至关重要。

### 7.1 明确实验目的

做实验之前一定要清楚实验的目的是什么，而不是盲目地做实验。一些要点能够帮助明确实验目的。

- **列举关心的科学问题**：明确做实验到底要验证什么结论或回答什么问题。
- **明确文章的创新点是什么**：所有的实验都是为了支持和解释这个创新点。
- **明确新的实验与已有实验的区别**：原则是一次只验证一个结论，不要调整多个变量。举个例子，现有模型是A和B结合，要验证一个新的策略，可以A保持不变而把B改成C，那么实验结果就能够说明C与B哪一个更优。如果同时将A改成了D，即使实验结果提高了也说不清提高到底是来自于对A的修改还是对B的修改。

### 7.2 实验数据构建

#### 7.2.1 公开数据

如果研究的问题有公开数据，这是最省事的情况。前面在讲实现Baseline的部分已经有讨论。在公开数据上进行实验的好处是便于与已有工作比较。挑战是竞争会非常激烈，因为大多数的人都追求省事，倾向于使用别人构建好的数据集。

#### 7.2.2 构建新数据

如果研究一个新问题或者新任务，需要自己构建数据集。此时的关键是要确保实验数据的可靠性，否则一切实验结论都是没有意义的。

为了训练模型和评价模型的能力，构建数据集需要人工进行标注，该过程中应该注意以下方面问题：

- **规模**：数据集的规模显然越大越好，但是更加费时费力，可以观察模型随训练数据量增加的变化，当实验结果趋于平稳的时候可以停止数据标注。多数情况需要折中或采取适当的方法，如利用众包进行多人标注。
- **规范**：标注者根据哪些原则进行标注？必须有基本的标注规范作为指导，而不能凭感觉。

- **质量**：判断标注数据的质量的一个方法是多个人背靠背地独立标注相同数据，计算多人标注之间的一致性（如计算Kappa值）。如果一致性较高，说明数据标注较可靠；如果一致性低，说明连人都无法达成共识，又怎么能要求机器做的准呢？如果训练数据里噪声多，无法训练出有效模型；如果测试数据中噪声多，难以客观、公平地验证模型的能力。当多人标注一致率不高的时候应该反复讨论修改标注规范，直至能够达到可以接受的一致率。

如果数据集构建是论文主要贡献之一，论文中应该完整地描述人工标注过程中的关键步骤和一致性评估等关键结果，否则一定会被拒掉。

## 7.3 性能评价

### 7.3.1 性能评价的多样性

更准、更快、更小是系统追求的目标。“准”即系统预测与预期的一致性高；“快”即训练或推断的速度快，其中推断速度对于系统应用来说更重要；“小”即在准度和速度相当的情况下，模型越小越好。准确性一般用精确率（accuracy），准确率（precision）、召回率（recall）和F值等评价指标来评估。通过阅读已有文献，可以很容易掌握目标任务常采用的评价指标。

很少有系统能够同时满足以上三点。因此，提炼创新点或者展示实验结果的时候要突出系统的优势。无论是在以上哪一个方面有提高，都算是有贡献。

## 7.4 基线系统

见前文提及的Baseline的选择和构建。

## 7.5 实验设置

所有对比系统都应该使用相同的实验设置。

测试集一定不能泄露，训练的时候是绝对不能看到测试集的。

神经网络模型对调参的要求非常高。要熟悉常用的调参技巧。做实验的时候要多尝试不同的参数组合（如优化函数、学习率、迭代次数、dropout比率等），增强对实验数据的感受。

## 7.6 实验结果展示与分析

研究的目的是不仅是实现一个方法，更重要的是给出结论，需要说明提出的方法是否有效，有多有效，在什么情况下有效，有什么局限。因此，实验结果的展示与分析非常重要。一个完整的实验至少要展示以下实验结果。

### 7.6.1 整体结果

提出方法的最终模型与Baseline的定量比较。这里要回答的问题是提出的方法是否优于已有方法，提高了多少，提高是否显著。

### 7.6.2 消融（Ablation）实验

消融实验的目的是为了解释一个系统的不同模块对最终模型的贡献。例如，一个基于特征的文本分类模型可能采用了多个类型的特征。那么哪些特征最有效呢？此类问题就可以通过消融实验来验证。

消融实验通常由两种方式。第一种是在一个Baseline的基础上逐一增加不同的模块（如不同类型的特征），增加哪个特征带来的提升大说明哪个特征比较有效。第二种是在最终模型上逐一去掉一个模块，去掉哪个模块导致系统性能的下降大，说明哪个模块比较有效。

消融实验也是一种定量分析。

### 7.6.3 定性分析

除了定量分析，定性分析有利于提高提出方法的可解释性和论文的可读性。定性分析主要通过一些实例直观地展示给读者。

一类定性分析是优势分析，即通过实例展示为什么提出的新方法比已有方法表现更好，使读者更加容易理解和接受提出的方法。



另一类定性分析是错误分析，对方法的预测错误进行分析和归纳，指出方法在处理哪些情况时会有困难，存在怎样的局限性。错误分析会给读者留下非常好的印象。一方面说明研究很深入，另一方面说明对方法的介绍很客观，并非一味吹嘘，而且指出了需要继续努力的地方。

## 8 写作

用一段文字来介绍科研写作是不现实的。如何写作值得独立文章甚至一本书来探讨。这里概括几个要点。

- 首要原则是**读者优先**。写文章的目的是交流思想，希望他人理解和接受提出的方法。很多读者不是同领域小同行，不能设想他们对相关概念和方法都了解。论文写作的大忌就是作者以为自己知道的东西，读者也应该知道。如果读者无法理解，那就是作者写作的失败。
- 第二原则是**逻辑**，不仅仅是写论文，包括实验和作报告！就像讲一个故事，前后呼应、连贯、有问有答、一一对应。思想→方案→结果→结论，环环相扣。
- 其余的才是所谓的一些技巧或经验。
  - 科研写作是八股文，有特定的格式和套路。
  - 一篇文章只要一个创新点，不要试图把所有东西都塞进去，提炼最精华的部分。
  - 用最简练的语言，多用简单句，少用复句。从模仿入手，总结常用的句式和结构。
  - 文字、图表规范，避免任何语法、格式等低级错误。
  - 用Latex写，不要用Word。
  - 参考第4.2节讨论如何审论文的部分，思考自己的文章能够有力回答审稿人关注的问题吗？

此外，除了写作，做报告也是表达能力的一种体现，需要提高这种意识，多加实践。

## 9 结论

本文简要地介绍了研究生科研中的若干关键问题，用几个关键词概括：**勤奋、主动、逻辑、细节、表达**。做研究既简单、又复杂，简单的目的、复杂的过程。这篇文章显然无法完全概括其中的所有内容，希望它能够给同学们提供一些基本的指导和帮助。

清华的唐杰老师提出从学生到学者要经历四个阶段：阶段一，在引路人的带领下，把一件事情做到最好；阶段二，想一个idea，在被告知哪些不能做后，把能做的部分做到极致；阶段三，完全独立地想一个idea，并独立地做完、做好；阶段四，能够带着第一阶段的人，引导他把一件事情做好。我认为很有道理。实验室希望培养出来的学生也能具备博士生的基本素质，特别是有些同学希望未来能够继续深造，甚至以后想在高校做老师，那么从现在开始努力。

做研究是一个没有止境的过程。我也一直在学习和领悟。这个文档还是一个草稿，希望同学们能够指出其中不清楚的部分或提出新的问题，帮助我进一步完善它。本文的很多内容都受到参考文献提及的优秀报告的影响，推荐大家去看一看这些资料，汲取其中的养料。

我要求同学们尽全力实践本文内容，包括但不限于：每人维护一个已读过论文的列表，写研究综述，提交周报或者进行汇报的时候要对照本文提及的内容，例如当展示实验结果的时候必须说明实验的设置并提供更详尽的定量和定性分析。

做研究真的挺难的，但也是很有趣的，有点像玩游戏打怪升级。不要畏惧困难，行动上紧张、思想上放松，寻找其中的乐趣，期待满满的成就感！加油哦：-)

## 参考文献

刘洋，浅谈研究生学位论文选题方法. (<https://www.bilibili.com/video/av94614779/>)

沈向洋，How to do research? (清华大学讲座<https://www.bilibili.com/s/video/BV1kk4y1C7xJ>)

沈向洋，You are how you read (<https://www.bilibili.com/s/video/BV1df4y1m74k>)

知乎问答：你有哪些深度学习调参技巧 (<https://www.zhihu.com/question/41631631>)

Richard Hamming. You and Your Research.

Chapman, David. How to do Research At the MIT AI Lab.