如何做研究，如何写论文

周志华

http://cs.nju.edu.cn/people/zhouzh/

2005年9月

为什么要做研究？

- 研究≠研发
  - 研究的目的是发现新知识、发明新技术
  - 研发：基于已有的知识和技术进行研制、开发
- 科学研究扩展人类的知识
- 没有科学研究就没有技术进步

为什么要写论文?

- 把你的工作告诉同行
  - 经过同行评审（peer-review），成为科学文献
- 基础研究的主要成果
  - 基础研究通常离实际应用有较大距离
  - 只有很少的研究工作能很快进入实际应用

前沿研究-> 实验室成熟技术/工业界新技术 -> 工业界成熟技术

论文好写吗?

- 很容易!

  只要有了好的研究工作，写论文不过就是用文字把你的工作描述出来

- 很难!

  如果没有研究工作支撑的话

论文是“做”出来的而不是“写”出来的

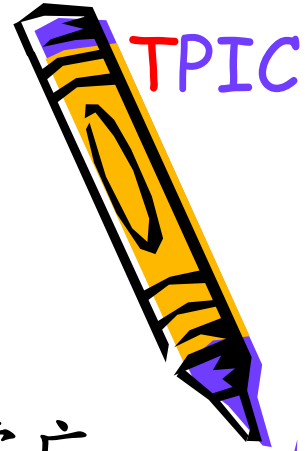
“写”的时间其实最多只占10%

如何做研究？

研究活动的大致过程- TPIC：

Topic -> Problem -> Idea -> Concrete work (theoretical analysis, experiments, etc.)

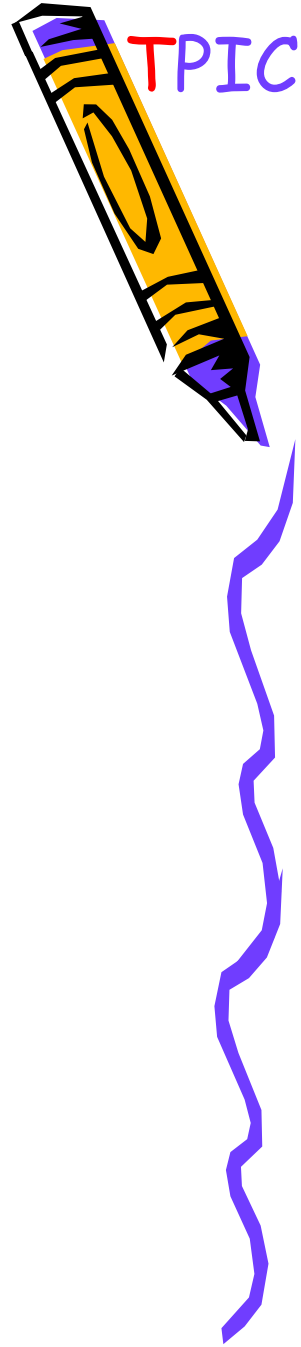
-> Paper writing -> Submit



从Topic开始！

- 计算机科学发展到今天，已经是一个非常广袤的学科
- 先要进入一个具体的分支学科和领域，并获得必要的了解

A rough taxonomy

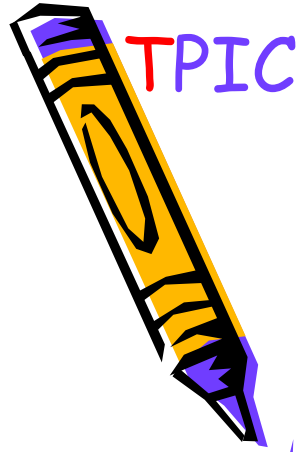


- **Computer Science**
  - Artificial Intelligence
    - Automated Reasoning (?)
    - Computer Vision (?)
    - Knowledge Engineering (?)
    - Knowledge Representation (?)
    - Machine Learning (?)
    - Multi-agent System (?)
    - Natural Language Processing (?)
    - Pattern Recognition (?)
    - Planning (?)
    - … … (?)
  - Computer Graphics
  - Database
  - Hardware
  - Multimedia
  - Network
  - Software Engineering
  - Theoretical Computer Science
  - … …
- **Physics**
- **… …**

即使在一个分支学科和领域中，也有太多的话题

For example

- Machine Learning
  - Active learning (?)
  - Clustering (?)
  - Decision tree (?)
  - Ensemble learning (?)
  - Incremental learning (?)
  - Inductive logic programming (?)
  - Lazy learning (?)
  - Multi-instance learning (?)
  - Multi-label learning (?)
  - Multi-strategy learning (?)
  - Neural network (?)
  - Regression (?)
  - Reinforcement learning (?)
  - Relational learning (?)
  - Rule induction (?)
  - Semi-supervised learning (?)
  - … … (?)



How ?

- 通常情况：导师给你一个topic
  - 导师往往是该领域的资深学者，对topic可能有较好的把握能力
  - 研究领域不存在“好”、“坏”之分，只要做得足够深入，都能做出好的工作
  - 在特定的时期，某些领域可能更活跃，相对来说杰出成果出得更多

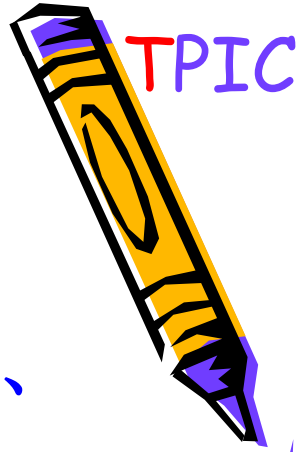
How (con't)?

- 然后，阅读关于该话题的重要文献，了解该话题的研究历程、研究现状
  - 请导师或该领域资深学者推荐读物
  - 基于导师推荐的读物，顺藤摸瓜（例如从参考文献）找到尽可能多的重要文献

读不懂的先跳过去，多读几遍

每个topic发展的历程都是无数聪明人智慧和汗水铺就的，这里面会有一条前后衔接的线索

某一天当你突然把头脑中杂乱的东西串成一线，你会觉得顿时豁然开朗："啊，原来是这样！"，这时你就开始尝到研究带来的乐趣了

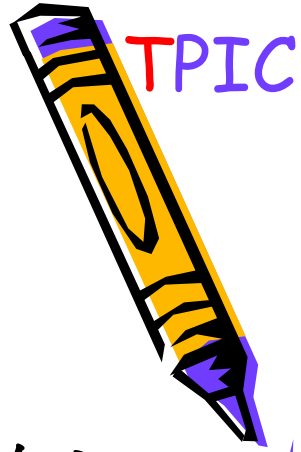


How (con't)?

- 如果因为种种原因，你需要自己去找topic、自己找东西读，那该怎么办?

办法1： 找人请教（自己要加以判断）

- 师兄、师姐
- 该领域的著名学者
  - 一定要有礼貌
  - 没有回音也不要难过
    - 名人可能每天会收到几十封类似你这样的信，不可能都回复
    - 名人自己可能有一群学生嗷嗷待哺
- 网络： BBS, mailing list, etc.



How (con't)?

办法2： 自力更生

① 搞清楚自己的领域里最重要的刊物、会议

问人、网上搜索、BBS, etc.

② 找来那上面最近几年的文章

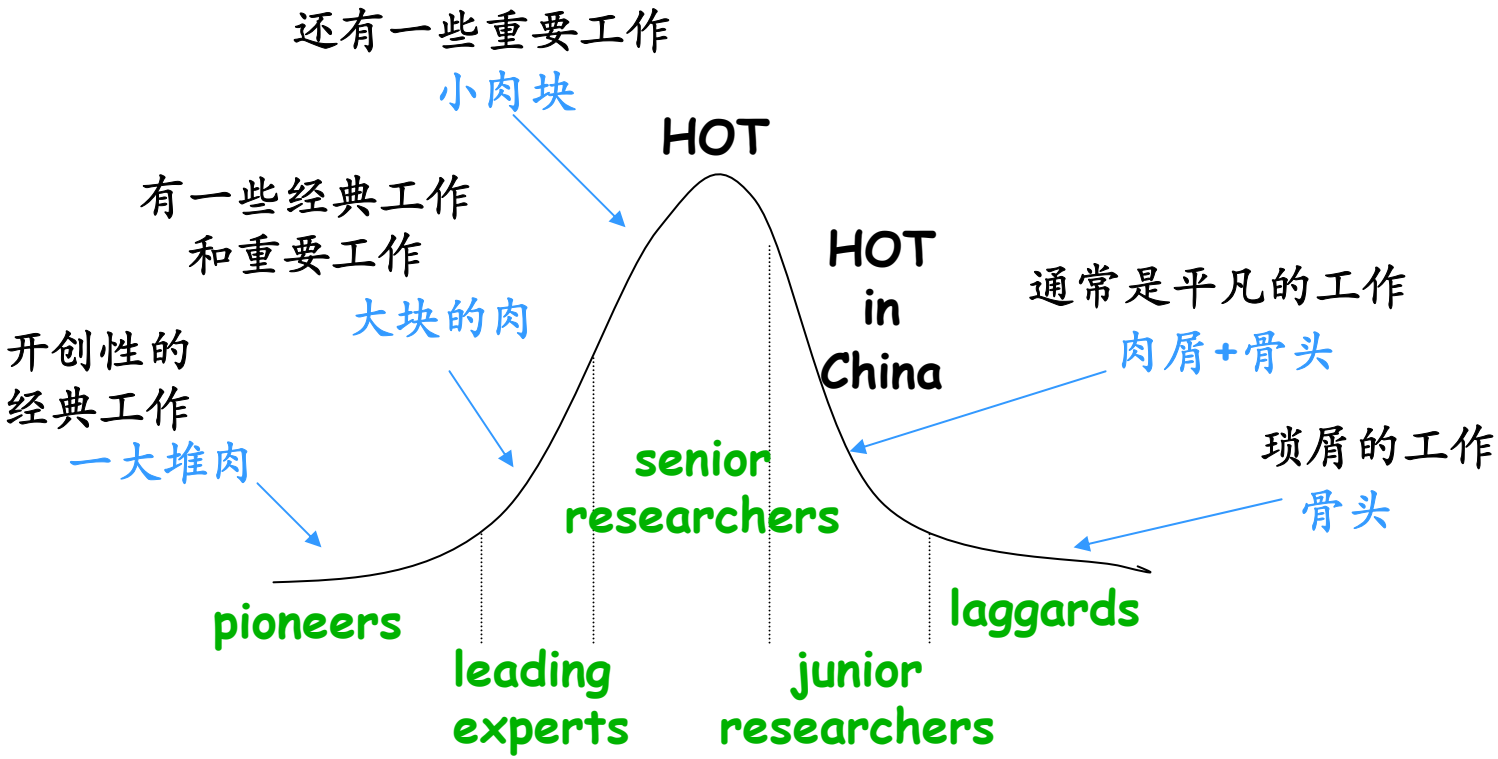
NJU数字图书馆、系图书馆、Internet, etc.

③读!

目的是为了大概知道有哪些topic，读摘要就可以了

可能要花很多时间

HOT - 热点



注意：

- 这只是目前CS中的一般情况，国内往往滞后一段时间
- 科学发展往往是螺旋式上升的："冷"了很久的一个topic，可能由于新的经典工作而重新"热"起来

How (con't)?

选择最适合你的topic:

① 自己的兴趣

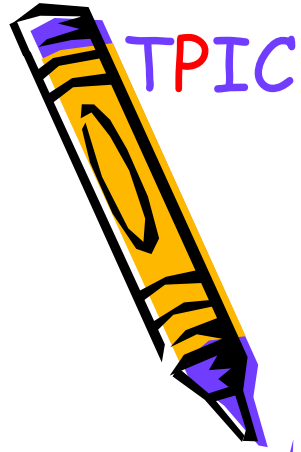
这是最重要的!

② 自己的知识结构

没有必要的知识积累，一切从头开始的话，……

③ 能否获得必要的资源

例如数据



Problem

“问题”是科学研究的心脏！

任何有价值的研究，都是为了解决某个问题

提出一个好的问题，已经成功了一半

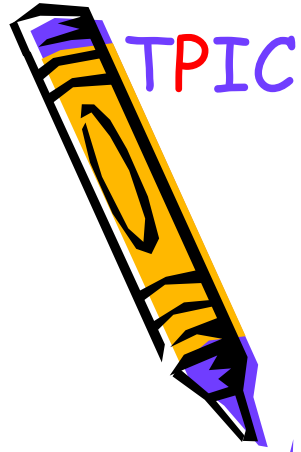
“问题”其实才是研究的真正开始

这可能是CS研究中最困难的部分

会找问题，是具有独立研究能力的标志

How ?

- 导师给你一个问题
  - 你很幸运!
  - 赶紧做，说不定导师忍不住自己去做了
- 自己产生问题，导师帮你判断
  - 这是通常的情况
  - 导师的研究经验和见解，会使得你少做无用功
- 有了问题以后：
  - 搞清楚该问题上已经有过的所有工作
    - 导师给你推荐重要文献，然后自己顺藤摸瓜
  - 优点是什么？缺点是什么？为什么没有彻底解决问题?
  - 悟出已有工作的发展线索

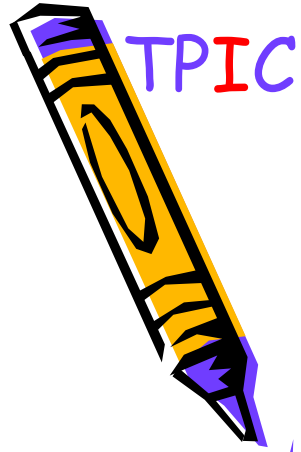


How ?

- 完全自力更生
  - 读发表在重要刊物和会议上的有关你的topic的文献
    - 2、30篇读下来，你大概能知道有哪些问题是没解决的了
    - 关注这个topic上活跃的leading expert的工作、他们的文章中可能会指出一些需要解决的重要问题
  - 兴趣
  - 有价值的问题
  - 知识结构
  - 资源
  - 宜“小题大做”，忌“大题小做”

对研究能力极大的锻炼

- 你可能会走很多弯路，有可能一无所成，但也有可能练就一身硬功夫



Idea

这是考验你聪明才智的时候了！

不要指望导师给你idea！

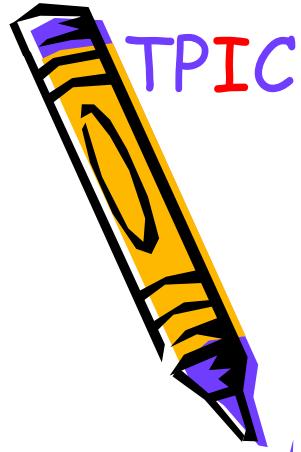
如果导师连idea都给你，你的运气实在是好到家了

你也许会有很多idea，导师可以帮助你判断、改善

有了好的idea，问题就解决了一大半

如果一下想不到好的idea，不要着急

博士期间能做出一项重要的工作就已经很好了



How ?

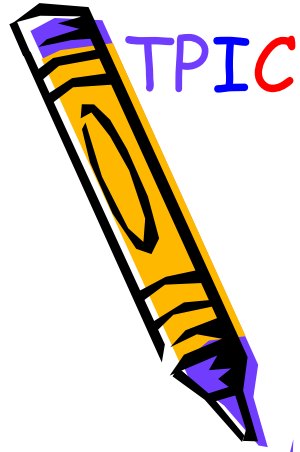
- 没人帮你判断idea怎么办?

  自力更生:

  - 是新的idea吗?
  - 是不是有道理?

    不是瞎蒙出来的
  - 是否可行?
  - 把这个idea先放到一边，过一周再考虑一下:

    是否有更好的办法?
  - 如果答案都是“yes”，那就赶紧动手吧!



Concrete work

Idea需要得到支持

CS里面通常是理论分析和实验验证

理论分析往往需要较好的数学功底

- 没有怎么办:

  1.学!

  2.找人合作

实验验证需要较好的实验设计能力

- 没有怎么办:

  1.学!

  2.找人合作

两者都不容易

到博士念完，至少在某方面应该是得心应手了

How ?

- 理论分析
  - 周全、不要有漏洞
  - 尽可能简单的工具
- 实验验证
  - 实验方案周全仔细
  - 基准测试
  - 其他学者也能使用的数据
  - 不可缺少的比较
  - 实验是可重复的
- 必要的分析和解释
  - Strength/weakness
  - How/When strong?
  - Why strong?
  - How/When weak?
  - Why weak?

小结一下

Topic -> Problem -> Idea -> Concrete work

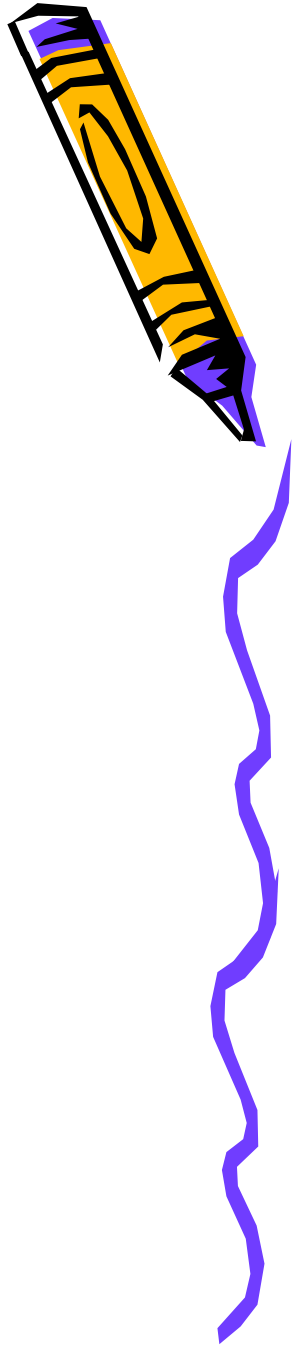
对一项具体的研究工作来说:

- 找到好的问题是非常重要的
- 有了好的idea，问题就解决了一大半
- 具体的工作是必不可少的

成功之路

聪明＋勤奋！

写论文之前



确定论文署名：

- 一般情况，按贡献大小排序
  - 如果是普通合作者，则考虑P, I, C
  - 如果是导师给问题，则主要考虑I, C
    - 导师是在培养学生
    - 综合考虑
    - 提出idea的人经常成为第一作者

- 通常由第一作者执笔
- 通讯作者通常是整个研究工作的负责人
- 作者不要太多，小的贡献可以放在致谢里

写论文之前

确定投稿目标：

- 根据工作的水准，挑选合适的发表源
  - 除非有特殊的考虑，否则：

    低投—遗憾，高投—延误发表
  - 针对不同的发表源可能有不同的写法
- 通常情况：导师帮你确定
- 如果需要自力更生，那么……

了解一下发表源

- 发表源的类型

Journal: 例如AIJ, IEEE Trans. PAMI

Conference: 例如IJCAI, AAAI

Magazine: 例如AI Mag, IEEE Intelligent System

计算机科学界很重视会议论文，高档次会议并不亚于高档次刊物

这是个例外，其他学科通常是不把会议当回事的

了解一下发表源 (con't)

- 发表源的档次
  - 由于CS发展迅速，很多分支学科都有好几个top journal和top conference

以AI为例：

Top journal: AIJ, TPAMI, MLJ, NCJ, IJCV

Leading journal: JAIR, CIJ, many IEEE Trans., etc.

Reputable journal: a lot

Top conference: IJCAI, AAAI, ICML, NIPS, CVPR, ICCV, UAI

Leading conference: ECCV, ECML, ECAI, etc.

Reputable conferences: a lot.

了解一下发表源 (con't)

- 发表源的档次

仅供参考:

- http://www.cs.iit.edu/~xli/CS-Journals-Rank.htm
- http://www.cc.gatech.edu/people/home/guofei/CS_ConfRank.htm
- http://woft.net (lily bbs) CompSci版精华区: 期刊、会议、投稿指南

国内刊物:

- 权威刊物: 中国科学，科学通报，自然科学进展
- 目前CS方面最国际化的刊物: JCST
- 一级学报：计算机学报，软件学报，电子学报，自动化学报
- 不错的刊物：计算机研究与发展，模式识别与人工智能，中文信息学报，……

了解一下发表源 (con't)

- 同一工作不能在刊物上发表两次
- 在会议上发表的工作，经过扩充后可以再投往刊物
  - 应该有至少30%的新东西（例如更详尽的理论分析、更多的实验结果等）
  - 投稿时根据刊物的要求明确说明
- CS刊物上文章篇幅通常比较长，会议文章篇幅较短
- CS刊物发表周期通常比较长，一些刊物可能要2-3年
- 绝不要一稿多投!

  现在有一些比较好的会议已经开始允许同时向刊物投稿

了解一下稿件处理流程

- Journal
  - 编辑部/主编(Editor-in-Chief)收到稿件
  - 转给合适的Associate Editor 处理
  - AE找reviewer审稿
  - Reviewer的意见到AE手里汇总
  - AE做出处理意见: accept, revision, reject
  - Editor-in-Chief审定

了解一下稿件处理流程(con't)

- 处理意见的大致类型:
  - Accept without revision
    - 罕见的情况
  - Minor revision
    - 比较少的情况，已经录用了
  - Major revision / Reject but resubmission encouraged
    - 多数情况；修改后再进行一轮审稿
  - Reject
    - 大多数文章，尤其是对很好的刊物来说
    - 很多刊物一期只有5、6篇文章

即使不能被录用，好刊物的审稿意见会对你有很大帮助

了解一下稿件处理流程

- Conference
  - 一般来说：
    - PC member对论文进行投标
    - PC Chair根据投标情况，把论文分配给适当的PC member
  - 有的会议设置了Area Chair，协助PC Chair处理某一领域的论文；有的会议没有投标过程

PC member 忙不过来时，往往会找external reviewer帮忙审稿

会议的处理意见通常是：录用或拒绝

文章类型

目前和各位有关的主要是：

- Journal
  - Regular paper
  - Short paper/Concise paper
  - Letter/Correspondence
- Conference
  - Oral
  - Poster

注意：

1. 不要自行与AE/Reviewer联系
2. 超过原定的审稿期后，可以写一封有礼貌的信给主编

审稿表

Example #1

Please review the attached manuscript with respect to:

- relevance
- significance,
- originality, and
- presentation.

Please indicate one of the following recommendations:

- accept unchanged,
- accept with minor changes,
- accept with *major changes*,
- reject but resubmission encouraged, or
- reject.

If changes are to be made, then please indicate them as *recommendations to the author (s)*; in case of rejection please give reasons for your decision, which can be communicated to the author(s).

Also indicate the extent to which the reviewed paper is within your area of expertise:

- fully,
- well,
- partially, or
- hardly.

审稿表

Organization
(Poor) (_)1 (_)2 (X)3 (_)4 (_)5 (Excellent)

Clarity
(Low) (_)1 (_)2 (X)3 (_)4 (_)5 (High)

Length
(Too Long) (_)1 (X)2 (_)3 (_)4 (_)5 (Too Short)

References
(Incomplete) (_)1 (X)2 (_)3 (_)4 (_)5 (Adequate)

Correctness
(Incorrect) (_)1 (_)2 (X)3 (_)4 (_)5 (Correct)

Significance
(Low) (_)1 (_)2 (X)3 (_)4 (_)5 (High)

Originality
(Low) (_)1 (_)2 (X)3 (_)4 (_)5 (High)

Attachments
(Unnecessary) (_)1 (_)2 (_)3 (X)4 (_)5 (Helpful)

If Survey Coverage
(Shallow) (_)1 (_)2 (_)3 (_)4 (_)5 (Broad)

Contribution
(No New Results) (_)1 (_)2 (X)3 (_)4 (_)5 (Significant)

*What are the contributions of the paper?

*What are the additional ways in which the paper could be improved?

Example #2

审稿表

1. How RELEVANT is this paper to AI researchers? (mark one box)
[ ] Very relevant [ ] Moderately relevant [ ] Not relevant

2. How SIGNIFICANT is this paper?
[ ] Very significant [ ] Moderately significant [ ] Not significant

3. How ORIGINAL is this paper?
[ ] Very original [ ] Moderately original [ ] Not original

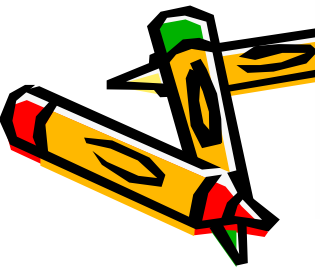
4. Is this paper technically SOUND?
[ ] Yes [ ] Seems valid, but did not check completely
[ ] Has minor errors [ ] Has major errors

5. How well is this paper PRESENTED?

	Good	Average	Poor
Overall organization	[ ]	[ ]	[ ]
English	[ ]	[ ]	[ ]
Readability	[ ]	[ ]	[ ]

6. Further comments, advice or explanations (Please be specific and constructive, especially with respect to any negative judgements above. Point to the section(s) where an error occurs, cite omitted references, etc.) Use as much space as you need.

Example #3



審稿表

Summary Report

Quality

- Outstanding (Paper Award Candidate) ☐
- Significant ☐
- Moderately significant ☐
- Marginal ☐
- Not significant ☐
- Erroneous or Trivial ☐

Recommendation: ★ Regular Paper ★ Correspondence ★ Note

- Accept without revision ☐
- Accept with minor revision ☐
- Review again after major revision ☐
- Resubmit after major revision ☐
- Submit to another journal ☐
- Reject ☐

Priority of Publication

- If accepted, should the manuscript be prioritized for publication? **Yes No**

Overview Report

(5: outstanding, 4: good, 3: fair, 2: poor, 1: very poor)

Content	**5-mark choices**				
Is the work relevant?	1	2	3	4	5
Is the work original?	1	2	3	4	5
Are the methods/proofs/experiments/etc. sound and convincing?	1	2	3	4	5
Presentation					
Is the abstract an adequate summary of the work?	1	2	3	4	5
Are the background and related work(s) clearly introduced?	1	2	3	4	5
Are the methods/proofs/experiments/etc. properly stated?	1	2	3	4	5
Are the conclusions clear and adequate?	1	2	3	4	5
Are the references adequate?	1	2	3	4	5
Is the presentation clear to the relevant audience?	1	2	3	4	5
Are the overall organization and length of the manuscript adequate?	1	2	3	4	5
Is the English satisfactory?	1	2	3	4	5

Example #4

审稿表

Example #5

题　目						编　号	
对本稿件总评价	创新性	很强　较强　一般　较差			选题意义	前沿　热门　一般	
	学术价值	很高　较高　一般　较低			应用价值	很高　较高　一般　较低	
	文字质量	优　良　中　差			文献引用	较全　不够全　缺关键性文献	
	综合学术水平	优（国际）	良	中		差	有重大错误
	录取意见	提前发表	发　表	修改后复审		退　稿	改投它刊
	发表形式	研究论文	技术报告	综　论		短　文	学术通讯

请在下面空白处提出具体意见：

1．稿件的创造性表现在哪里？
2．稿件所述内容是否代表当今发展方向？
3．在稿件所研究的领域中，热点话题是什么？
4．稿件中有无不妥之处，并提出具体修改意见。
5．稿件需要作退稿处理，请说明理由。

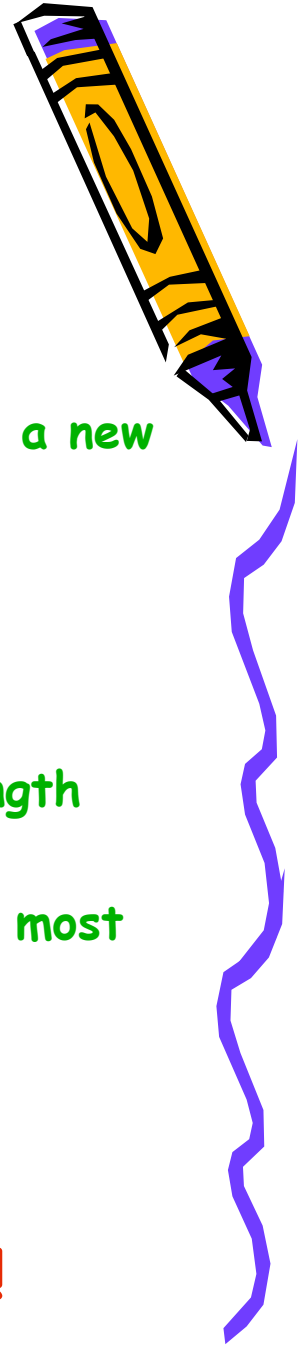
审稿人单位：
邮政编码：
电话（或 Email）：

审稿人（签名）
年　月　日

Reviewer如何审稿?

- 一般来说，主要考察以下方面：
  - Relevance：是否相关
  - Originality：是否有创新
  - Significance：是否重要
  - Soundness：是否有足够强的支持
  - Presentation：是否表达得好

Reviewer如何审稿 (con't)?



In summary, reviewers will check:

- Does the paper introduce a new problem or provide a new solution to an existing one?
- What is the main result of this paper?
- Is the result significant?
- Is the paper technically sound?
- Does the paper provide an assessment of the strength and weakness of the results?
- Is the paper clearly written so as to accessible by most researchers in this area?
- Does the paper refer appropriate related works?
- … …

Don't attempt to fool the reviewers !

如何写论文？

你需要说的其实就是这些：

- Problem X is important
- Previous works A, B, and C have been done
- A, B, and C have their weakness
- Your work D
- Theoretic analysis
- Experimental comparison against A, B, and C
- Why D is better
- Strength and weakness of D
- Future works on D

Outline

科学论文有比较清晰的结构，一般包括：

- Title
- Abstract
- Introduction
- Previous works
- Your contribution
- Support (theoretical or experimental)
- Discussion
- Conclusion
- Reference

Title

- 清楚地表达出主要工作
- 字数忌长（尽可能不要超过20字）
- 有吸引力

An example:

“甲地区乙时期丙昆虫交配过程的跟踪研究及其结果”

“昆虫交配后吃掉配偶的原因探讨”

“昆虫求偶过程中的献身行为”

“Eat me!”

The title of a <Science> paper

Abstract

Purpose: Summarize your contributions

Style:

- What is the problem
- What is your work
- Features of your work
- Advantages of your work
- Results

Introduction

Purpose: Background and organization of the paper

Style:

- Problem X is important
- A, B, and C have been done
- A, B, and C have their weakness
- Our work D
- Features and advantages of D
- Results
- Organization of the paper

Previous works

Purpose: Why your work, the differences

Style:

- Categorization of previous works
- One or two sentences for a work
  - Strength
  - Weakness

Don't over-criticize previous works

Your contribution

Purpose: Introduce your work

Style:

- **Motivation**
- **Definition, notation**
- **Algorithm**
  - **Pseudo-code**
  - **Diagram**
  - **Explanations**

If you were the reader, what questions you want to ask?

Theoretical analysis

Purpose: Theoretical support to your work

Style:

- Definition, notation (can be omitted if existing)
- Lemma
- Theorem
- Proof

Put tedious details in Appendix

Experiments

Purpose: Experimental support to your work

Style:

- **Experimental design**

  Be sure that other researchers can repeat your experiments according to your descriptions

- **Comparisons**

- **Discussions**

  What is revealed by the experiments?

Put tedious details in Appendix

Discussion

Purpose: The relationship between your work and some very related works

Style:

- Work A
  - Why it is very related
  - Difference to your work
- Work B
  - Why it is very related
  - Difference to your work

Conclusion

Purpose: Summary and future works

Style:

- Summary
- Future works

Acknowledgement

Reference

Appendix

Misconceptions 1:

The more, the better

Don't provide too many "new" ideas in a paper. Try your best to focus on your core work

Misconceptions 2:

The more complex, the better

Try to make your paper easy to be understood.

Remember: You are showing your thoughts instead of confusing the readers

Mathematical language is to help you clearly express your thoughts, not to show your depth

Misconceptions 3:

The more selling, the better

Don't exaggerate too much on your work.
The reviewers/readers will judge it.

Misconceptions 4:

The more authoritative, the better

Don't refer too many own works while ignoring others

Don't behave as a bigshot if you are not

Tricks

- Read more, write more
- Use good papers as samples
  - to simulate but not to copy
- Use simple sentences
- Not to show your elegance
  - unless you are very famous
- Try to reduce syntax errors
  - Good works won't be rejected for syntax, but at least the paper should be readable

Tricks (con't)

- Make your abbreviations easy to be pronounced
  - e.g. STARE is better than SBRE
    "STAtistics-based Rule Extraction"
- Typeset according to the requirements of the target
- Revise, revise, and revise
  - each time lay your paper aside for some time
  - ask other persons to read your paper
- Improve from comments

古今做学问者必经的三个境界：

- 昨夜西风凋碧树，独上高楼，
  望尽天涯路
- 衣带渐宽终不悔，为伊消得人憔悴
- 众里寻他千百度，蓦然回首，
  那人却在灯火阑珊处

学问苦乐，日久自悟

王国维先生之言，录与各位共勉

希望能对各位有点小帮助，

谢谢！