

2021 年论文写作与学术规范复习归纳

考试时间：2021-7-2 2:00 PM-4:00 PM

考试地点：A2-307 A2-308

目录

1. 如何选择主题 (topic)	1
2. 开始.....	2
3. 结构和组织.....	3
4. 引用.....	5
5. 摘要和标题.....	7
6. 图表.....	8
7. 作者身份 (Authorship)	15
8. 双重出版.....	16
9. 抄袭.....	17
10. 选择要发表的期刊.....	17
11. 编辑看什么.....	18
12. 推荐信 (cover letter)	19
13. 编辑审查过程.....	19
14. 综述 (review)	22
15. 科学发表的伦理问题.....	23
16. 编辑的伦理准则.....	25
17. (对学生的) 一些建议.....	25
附件一：介绍的修改样例.....	25
附件二：实验和分析的示例.....	30
附件三：结论和未来的工作的示例.....	31
附件四：一个摘要修改的例子.....	32
附件五：一个同行评议的例子.....	33

1. 如何选择主题 (topic)

在写论文的过程中，没有比选择主题更重要的一步。你像一个选择去哪的旅行者。如果是你喜欢的地方，你会喜欢去。如果换成不喜欢的地方，你去那里会很痛苦。理想情况下，你应该选一个感兴趣的课题，足够复杂，需要几个研究来源，且不会让读者厌烦或有居高临下的感觉。

1.1 如何选择研究主题

(1) 选一个喜欢的主题。无论这个主题是历史的、有争议的 (controversial) 或字面意义上的，你的选择必须满足 2 个需求：它必须被你的上级批准；而且，更重要的是，它必须对你有吸引力。

(2) 如果你对这个主题完全迷茫，并且绞尽脑汁也想不出该写什么才能写 10 页，去图书馆浏览。

- (3) 你的图书管理员能引导你到其他电子存储资源。
- (4) 如果你依然不能找到一个合适的主题，是时候打开电脑并在互联网上搜索想法了。
- (5) 百科全书也是一个关于可能的主题的不可思议的富矿。
- (6) 花时间搜索主题。不要坚持选定脑海里的第一个想法。仔细思考。问自己是否会喜欢花 5 周时间来阅读关于这个主题的材料并写作关于这个主题的文章。如果有困惑，一直看，直到碰到让你兴奋的主意。
- (7) 找出主题并缩小范围的最好方法之一是问自己关于自己的一般兴趣的问题。你可能通过问自己一些一般的问题来开始：我想写什么？我对什么样的特定主题感兴趣？我真正喜欢什么？如果你不能马上回答就一直问。一旦你有了答案，用它来问另一个目标更明确的问题。

1.2 要避免的主题

- (1) 太大的主题：解决方案要缩小到让文章不平凡的主题。
- (2) 基于一个来源的主题：只选择足够宽广有多个来源的、不被一个作者的观点左右的主题。
- (3) 不重要的（trivial）主题：你的判断必须避开平凡的主题。最安全也最明智的选择：也是问你的主管。
- (4) 太热门的主题：最好避免在公共领域仍然有火药味的争论的主题。

1.3 缩小主题的范围

没有巨蟒知道它的嘴的维度，但是所有巨蟒都本能地知道它不能吞下一个大象。牢记文章的长度与主题的对比。大多数研究论文是 10 页左右——一些书有更长的前言。

2. 开始

2.1 为何要写论文、发论文？

如果你想写一篇科学论文并经过同行评议（peer review）发表，你的第一步应该是什么？你应该问自己写科学论文的动机。

- (1) 写经过同行评议而出版的论文是许多科学家和工程师职业生涯中重要的一部分。它也是科学事业中必不可少的一部分。
- (2) 在经过同行评议的期刊上写论文并发表艰难的工作。
- (3) 为何人们要这样做？
- (4) 什么激励作者通过这样的写作过程和同行评议过程，以发表他们的工作？

两种动机：

利他主义（altruism）：在今天，经过同行评议的科学出版物是传播和获得科学前沿信息的主要方法。大部分科学家有一种强烈的欲望来为他们的领域前沿做贡献，这常常是他们当科学家的主要原因。出版物通常是做贡献最直接的方式。

利己主义（self-interest）：发文章也能给作者带来实际的利益，因此给写作并发表文章提供了利己主义的动机。发文章可能是职业生涯发展的需要，并常常伴随着直接或间接的金钱奖励。

平衡利他主义和利己主义：

大多数作者同时有这两种动机。当利他主义和利己主义不平衡的时候问题就会出现。尤其是，如果利己主义太强，以至于变成自私并淹没了科学进步的利他目标，整个科学事业就会遇到问题。

2.2 资料搜索

一个新的研究项目几乎总是从资料搜索开始。搜索的目标是，在开始增加共同体知识的探索前，估计这个主题的共同体知识的状态。因为科学是关于确认或否认现存知识或者发展新知识的，所以对共同体知识当前状态的彻底了解是必要的。此外，资料搜索会成为引证的基础。

注意，资料搜索不是寻找相关论文，它是阅读相关论文。

5 个改善资料搜索的提示：

- (1) 在研究前进行资料搜索，至少必须在写论文前。
- (2) 接下来，在你找到的相关论文中，大多数有希望并被阅读的论文常常是被参考的论文。
- (3) 查找你指定之外的领域（Look in fields outside your discipline）【这常常意味着查找不同的搜索关键词，而这些关键词是在阅读范围外资料时发现的】。
- (4) 当你关于之前值得引用的论文的记忆是一个好的开始时，没人知道资料的全貌，甚至最小的 niche 领域中的资料也不行。不要只依赖你的记忆。
- (5) 当完成稿子时，查找该主题最近的出版物。其他研究人员经常在相似的主题下工作，而且可能已经发表了论文，所以应该阅读这些论文以保证你的稿子包含领域内最新的共同体知识。

2.3 把发表放在心上来计划并执行研究

大部分项目以产出论文为目的，或者至少有这种可能性的想法。如果是，应该把发表放在心上来计划并执行研究。

一个对科学论文的关键需求是足够细节地记载工作，使得读者能跟上所展示的推理并确认所得到的结论。

进一步讲，已发表论文的作者必须愿意面对批评并为自己的工作辩护，所以他们应该有可用的原始数据和关于实验过程的重要细节以供之后的审查。

3. 结构和组织

一个标准结构的例子，来自于以下期刊论文：

Zisen Fang, Xiaowei Yang, Le Han, and Xiaolan Liu, A sequentially truncated higher singular value decomposition based algorithm for tensor completion, IEEE Transactions on Cybernetics, 2019, 49(5): 1956-1967.

3.1 科学论文的标准结构

今天，大部分发表在科学期刊上的论文遵循一个相当简单的结构。包括一些变体，大多数论文使用“IMRaD”格式：

介绍（Introduction）

方法（Method）【实验、理论、设计、模型】

结果和讨论（Results and Discussion）

结论（Conclusions）

使用 IMRaD 结构有 2 个主要的优点：它让作者更容易组织论文的内容，并让读者更容易适时地（opportunisticly）找到他们想要的信息。接下来的章节可以看作这些标准章节的细化。

3.2 介绍

一篇论文在发表在科学期刊前必须达到 4 个标准：

- (1) 论文的内容必须符合期刊的范围。
- (2) 论文的质量（研究的方法和执行，以及写作）必须足够高。
- (3) 它必须展现新颖的结果（除了综述论文及其类似论文）。
- (4) 结果必须具有足够的意义且值得阅读（因此值得发表）。

在这 4 个标准中，作者应该在介绍中清晰地声称拥有 3 个（范围、新颖性和重要性）。

介绍的基本流程：

介绍的基本流程是从一般到特殊。

步骤 1：建立一个领域（工作的领域是什么，这个领域为何重要，做了什么？）。

步骤 2：建立一个商机（niche）【指出一个空白，提出一个问题，或挑战该领域之前的工作】。

步骤 3：占据这个商机（概述目的并宣布现在的研究；可以对结果做个总结）。

介绍的修改样例（[附件一](#)）。

3.3 方法

方法章节（有时被称为材料和方法【Materials and Method】章节）描述结果是怎样产生的。它应该被充分地详细描述，使得一个在相同领域的独立的研究人员能复制结果，充分验证结论的有效性。

通常情况下，这不需要清楚的一步一步的指令，但某种程度上对先前出版物的引用提供了这样的细节。对于一些研究文章，它的方法是新的。对于这些例子，一个更详细的描述是需要的。对于标准的或被确认的方法，指出方法的名字就足够了。

让我们稍微更认真地分析“充足的细节”的需求。在工作中有 2 种相互关联的目标：读者应该获得复制实验结果和判断结果的能力。

即使非常少读者试图复制另一个实验，大部分认真的读者也会试图判断他们所阅读工作的正确性。

内部的正确性意味着所得的结论被所展示的结果支持。外部的正确性则与结论能被泛化的程度相关。

没有认真写作的方法章节，就不可能判断工作的正确性。

一个“方法”比实验方法更广泛使用。

方法能包括一个理论的发展（无论是作为必要的背景还是作为论文的新元素）、一种特定设备设计的确立，或者是一种建模工具的开发或描述。

一种常见的 IMRaD 结构变体将理论（或设计，或建模）分成它自己的前导章节和后面的实验方法。

一个好的方法章节应该不经描述做了什么和怎样做的，还要解释实验设计。在许多可用的方法中，为何选择了这个方法？统计学上的考虑，比如所用的采样方法和分析方法，应该被解释。也要考虑插图或图表能如何用于说明或概述方法。

创新部分：

Adaptive Method for Low-Rank Approximation of a Tensor
Dealing with Incomplete Tensors
Convergence Analysis

Complexity Analysis

3.4 结果和讨论

论文的结果，如果作为独立的章节，应该非常短。它简单地展示了前面章节所述方法对应获得的结果，组织起来让它们对读者可得。通常，这些结果以表格和/或图的方式展示。从支持论文主体的文本的角度来看，精心制作的表格和插图需要非常小，所以结果通常是和它们的讨论合并成结果和讨论这一章节。

展示结果时，一个重要的目标是清楚地表明那些结果是新的，当正确地引用之前已被发表的结果时。

讨论：

讨论章节的目的是解释结果并展示它们如何帮助解答介绍中提出的研究问题。

这讨论一般经过概述结果、讨论结果是否理想、比较这些结果和之前的工作、说明并解释结果（常常与一个理论或模型比较）和假设它们的普遍性这几个步骤。

讨论章节倒置了介绍的格式，从特殊（在这项工作中产生的结果）到一般（这些结果如何说明一个能被更广泛应用的一般原则）。

在这项工作的过程中遇到的所有问题或缺点（**shortcoming**）也应该被讨论，尤其是，如果它们可能影响被如何诠释的。

实验和分析的示例（[附件二](#)）。

3.5 结论

结论章节提供了一个结果和讨论的简短摘要，但是它应该不只是一个摘要。在展示如何解决介绍中提出的研究问题后，这些发现的影响应该被强调，解释这项工作有多重要。这里的目标是提供能被证据支持的最一般的断言。

这一章节应该是最吸引读者的，避免列出所有“我”或“我们”完成的所有东西（**avoiding a list of all the things that “I” or “we” have accomplished**）。

结论章节应该允许投机阅读（**opportunistic reading**）。在写这章时，想象一个读者，读完了介绍，浏览了插图，然后跳到结论。结论应该简明地提供作者想传达的关键信息。它不应该重复结果和讨论中的论点，只给出最终和最一般的结论。

结果和讨论章节往往很长，与此相反，结论章节一般很短。

写结论的常见陷阱：

写结论时的一些常见的陷阱是重复摘要、重复介绍中的背景信息、介绍结果和讨论中未发现的新证据或新论点、重复结果和讨论中的论点，或无法解决介绍中提出的所有研究问题。

结论和未来的工作的示例（[附件三](#)）。

4. 引用

即使有许多形式的科学出版物，最常见的两种也就是会议报告（**conference presentation**）【有或没有一些非同评议的文本】和经过同行评议的期刊论文（无论是印刷的还是在线的）。

因为几乎所有的科学前沿都建立在之前的知识上，所以将新的工作放在适当的背景下，尊重它所基于的先前的工作就非常关键。为此，首要的机制就是引用（**citation, reference**）。

引用也被用来帮助读者确认新工作的质量并获得结论的强度（**strength**）。

4.1 引用的 5 个目标

引用的一般定义：

引用被定义为对一个信息源或数据源的提及（**reference**）。能被引用的东西包括期刊文章、会议论文（**conference proceedings**）、书籍、学位论文（**student theses**）、报纸、非印刷源（例如电影或其他记录媒介）、网站或其他在线资源、计算机材料（例如出版的数据或软件光盘），以及个人通信。

如何引用：

引用应该放在文本中，必须清楚地表明需要引用的材料。

显然，引用必须提供足够的细节，以至于被引用的材料能被找到并独一无二地标识。

所以，每个期刊建立了特定的格式，引用时必须遵守。（换句话说，没有所有期刊都遵守的统一格式。）

引用的 5 个目标：

（1）提供足够的工作背景，使得其他人能够进行这项工作的关键分析，因而使读者能为自己评估（**gauge**）作者的结论是否合理；

（2）给读者提供背景来源和相关材料，使得当前工作能被目标读者明白（因此创造一个科学网【**web**】）；

（3）为读者建立可信度（例如，作者知道这个领域，已经做了他们的功课，等等）并/或告知读者这篇论文属于一个特定的学派（**school of thought**）；

（4）提供关于用于比较的代替的想法、数据或结论的例子并与这项工作对比；

（5）为这项工作所依赖的源给出承认和信誉（也就是，确认使用了作者的想法或数据），因此维护知识上的正直（**intellectual honesty**）。

好的引用：

如果他们阅读了所有的引用，他们会有足够的背景知识来明白你的工作吗？会有任何引用是不必要的或冗余的（因此浪费读者的时间）吗？

一个简单的测试是非常好的机会，足以决定大多数引用是否合适：在此增加这个引用会让这篇论文对读者更有价值，还是相反？

4.2 文献检索

在开始研究之前，你应该对该领域的关键论文有一个很好的了解。这种文献搜索应该在你的研究期间更新，特别是当有新的想法或方向改变时。

另外，其他研究人员也经常从事类似的课题，并且可能在你完成原始文献搜索后发表了论文。

4.3 验证、验证、验证

引用最普遍的问题之一是它们经常是不完整或不准确的。验证参考文献的准确性是作者的工作。编辑、校对者和审稿人不对参考文献的准确性负责，也不被期望去检查参考文献的准确性。尽管文字编辑（**copyeditor**）试图标记不完整或格式不正确的引用，但最终必须由作者修复发现的错误。

为什么不预先做这些工作，以确保引用是完整的、准确的和正确格式化的？

验证：

然而，还有另一个问题可能也在起作用：复制其他论文的引用。换句话说，一些作者犯了严重引用错误，添加了一个没有读过的参考文献到论文中。

在没有真正查阅和阅读论文的情况下，复制其他论文的引用会导致错误的传播，这些错误永远不会得到纠正。

所有未读引用都应该避免：只引用你读过的论文。

4.4 引用的其他问题

(1) 虚假引用：不需要但无论如何都被包括的引用。

(2) 有偏引用：除了满足引用的五个目标之外，还添加（或省略）了引用。偏差包括过度引用朋友或同事的工作，省略对竞争对手工作的引用，和无目的引用。

(3) 自引：对自己工作的引用。

(4) 排除对立证据：一种有偏引用，以前的工作的结论或数据与当前的工作相悖的引用被省略。

5. 摘要和标题

标题和摘要的目的通常被描述为“销售”论文：让某人阅读标题以阅读摘要，而某人阅读摘要以进一步阅读该论文。

标题（紧随摘要之后）是读者首先看到的内容，因此它应该是作者撰写的最后一件事（紧随摘要之后）。因为摘要应该写在标题之前，所以我将首先讨论摘要。

5.1 撰写摘要

摘要应该是论文的简洁和单独的一个概括，应涵盖以下主题：

背景/动机/上下文

目的/目标/问题说明

方法/方法/程序/材料

结果

结论/含义

典型的摘要约为 150-200 个单词（尽管最大允许长度取决于期刊和论文类型），因此每个单词都必须谨慎选择。

5.2 结构化摘要

背景：什么问题导致了这项工作？在什么环境下使这项工作变得有趣或重要？

目的：您打算在这项工作中实现什么？正在填补什么空白？

方法：您是如何着手实现目标的（例如实验方法，模拟方法，理论方法，这些方法的组合等）？你实际在做什么？

结果：研究的主要结果是什么（包括适当的数字）？

结论：您的主要结论是什么？为什么结果很重要？这些结论产生了哪方面的影响？

5.3 关于摘要的重要附加注意事项

摘要完成后，应将摘要分为三部分进行测试：

(1) 摘要中的所有信息是否与论文正文中的内容一致？

(2) 摘要中找到的所有信息也可以在本文正文中找到吗？

(3) 论文的重要信息是否在摘要中找到？论文摘要中是否缺少关键词？

摘要必须是独立的（**self-contained**），并且通常不应包含对其他论文的引用。如果需要引用（例如，如果论文是对先前出版物的回应），则必须将完整的引用嵌入摘要中。

避免在摘要中使用缩写词（**abbreviation**）和首字母缩写词（**acronym**），但是如果您坚持要使用缩写词和首字母缩写词，请在首次使用时将其拼写清楚。

一个摘要修改的例子（[附件四](#)）。

5.4 标题

摘要一旦写完，就该写标题了。

手稿和摘要完成后，几乎肯定需要修改标题。

标题一般应反映工作的目的和方法。

与报纸报道和营销新闻稿的写法不同，科学论文的标题应该描述工作的目的，而不是结果。

例子：

原标题：A New Strategy based on ST-HOSVD for Tensor Factorization with Missing Data

修改后：A Sequentially Truncated Higher Order Singular Value Decomposition Based Algorithm for Tensor Completion

翻译：

原标题：基于 ST-HOSVD 的缺失数据张量分解的新策略

修改后：基于顺序截断的高阶奇异值分解的张量完成算法

5.5 关键词

确定关键字的重要方法很简单：如果您正在寻找有关论文主题的文章，那么您将在搜索引擎中键入哪些词才能找到它？

找到了不错的关键字列表后，请回头查看标题和摘要。是否在标题和摘要中找到了这些关键字？如果不是这样，搜索您的文章的人可能会很容易错过它。最重要的关键字应该在标题和摘要中找到几次。

6. 图表

图形是任何科学出版物中极其重要的一部分。因为几乎所有的科学出版物都包含了需要推理的定量信息，所以图形几乎总是必要的。

6.1 使用图形的目的

作为一种通信形式，图形（特别是定量数据的图形显示）特别适合于快速有效地传递来自复杂数据集的信息。

图形利用了人类大脑的强大功能来识别视觉/空间模式，并快速地将焦点从大图转移到小细节。

像所有形式的交流一样，图形可以用来解释和澄清，但也可以用来迷惑或欺骗。因此，图形的首要规则很简单：它们必须帮助揭示真相。

图表的目的是什么？以下是在科学出版物中使用图表进行交流的一些更重要的目标：

(1) 记录数据（图表通常是数据发布的唯一位置）；

- (2) 进行比较（如显示趋势）；
- (3) 考虑到原因和结果的推论；
- (4) 讲述一个故事，或者至少是一个组成部分（不可分割的部分）的故事；
- (5) 与正文相结合，加强论文的整体传播。

创建图形显示的过程：

总的来说，创建图形显示的过程遵循以下基本步骤：

- 选择要显示的数据
- 定义要传递的消息
- 选择支持该消息的图形样式
- 构造图以寻求清晰度
- 修改它直到它是正确的

6.2 图形的完整性（Integrity）

图表是一种强大的交流工具，人们必须选择交流真相而不是谎言。我建议这些问题作为一个测试图形完整性：

- 展览揭示了真相吗？
- 表述是否准确？
- 数据是否被仔细记录？
- 显示的方法能避免错误的数据读数吗？
- 是否显示了适当的比较和上下文？
- 您选择了要显示的正确数据了吗？
- 数据中的不确定性能否得到恰当的评估？
- 其他人能根据你提供的信息来验证你的结论吗？

6.3 一些指导原则

记住，一个数据有四个部分：描述（它是什么？），一个数字，一个单位，和一个不确定性估计。如果缺少这四项中的任何一项，那么这些数据本质上是无用的。

在绘制数据时，尝试将数据的所有四个部分放入图中。

如果有数据点被删除，解释原因。

如果存在误差条（它们几乎总是应该存在），清楚地解释它们代表什么（数据样本的一个标准差，均值的一个标准误差，一个具体的置信区间，等等）。

对于数据，上下文总是很重要，对于数据的显示也是如此。“图表不能断章取义地引用数据。”

让数据脱颖而出——不要让它迷失在杂乱的行和标签中。快速浏览应该允许您将每个数据点与图表上的其他数据点区分开来。

表格最适合用于查找特定的信息或确切的值，而图表则最适合用于显示趋势和进行比较。如果你认为读者会尝试从图表中读出数字，考虑一个表格（代替或附加）。

当数据点较少时，表通常比图更可取。

较高的数据密度是好的，只要不牺牲准确性和清晰度。

无论如何，在可以增强图像效果的时候使用彩色（现在大多数文章都是在线阅读的），但是要确保用黑白打印时不会丢失任何信息。

根据需要在图表或标题中添加标签，以尽量减少来回引用文本的需要。如果可能，这个数字应该是可解释的。

图形说明不应该是事后添加的——它们是图形不可分割的一部分。计划标题与图表一起显示上下文和数据的解释。同样，我们的目标是尽可能使图形本身具有可解释性。

理想情况下，图形说明将做三件事：描述图表中的所有内容，吸引人们对其重要特征的注意，（如果可行的话）描述从图中得出的主要结论。

图表不应该有标题。将标题信息放入图标题中。

确保图表的每个元素都得到了充分的解释，如果不是在图表或其标题中，那么就是在文本中。

饼状图几乎从来不是最好的选择。

只有当你找不到更好的选择时，才使用柱状图。柱状图只能用于绘制分类数据，但如果类别有一个自然顺序，那么线图通常会更好。因为杆的长度代表数量的大小，必须薄酒吧（酒吧区不混淆读者）和 y 轴必须从 0 开始（这个限制的原因之一，其他图表类型往往优先于柱状图）。

通常，并排条形比较起来比堆叠条形比较好，因为堆叠底部的起伏会使堆叠的上部难以解释。堆积的折线图也遭受这些相同的困难。

避免所有虚假的三维（3D）效果，例如在柱状图中使用 3D 条。它们只会导致混乱，而不会带来更大的清晰。

图应该尽可能简单，而且图不应该比它所表示的数据更复杂。

使用对数尺度来揭示数据中的趋势，而不是隐藏它们。对数尺度强调相对变化，而线性尺度最擅长显示绝对变化。

如果合适的话，可以考虑为每个轴使用两个刻度（例如，一个刻度显示实际值，另一个刻度显示该值与引用的百分比变化）。

数据聚合或减少（将数据分组并绘制组摘要）可以抑制干扰并揭示趋势，但只有在正确操作的情况下才可以。例如，直方图通常对容器大小和起始点非常敏感。时间序列图对选择的开始时间和间隔也很敏感。如果您根据任意选择的聚合参数得出关于数据更改的结论，那么要非常小心。

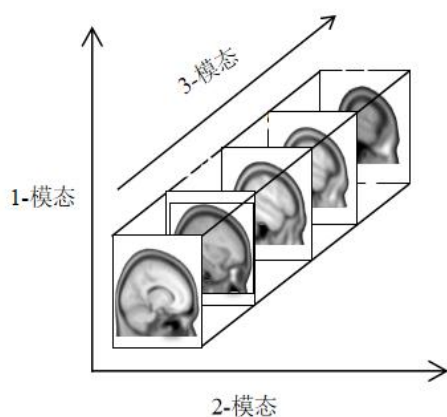
选择绘图比例（例如，x 和 y 轴的开始和停止值）以避免空白：尽量使用每个比例的至少 80% 来显示数据。

在进行比较时，基线有时是很重要的，但如果没有自然的基线，要注意任意的选择是如何将特定的解释推给读者的。0 可能是一个自然的基线，但是不要在绘图尺度上强制为 0，如果这样会导致浪费图形空间。

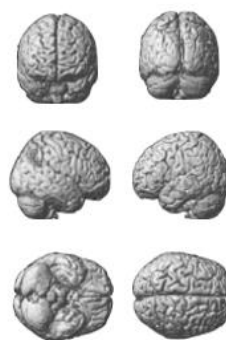
不要在单个图形的轴上使用缩放中断或改变缩放。如果需要两个刻度来显示数据，请使用两个图形（或尝试使用对数刻度以获得更好的分辨率）。

不能用一个好的图表来修正错误的的数据。

fMRI 影像脑部诊断数据



fMRI数据的三阶张量图解

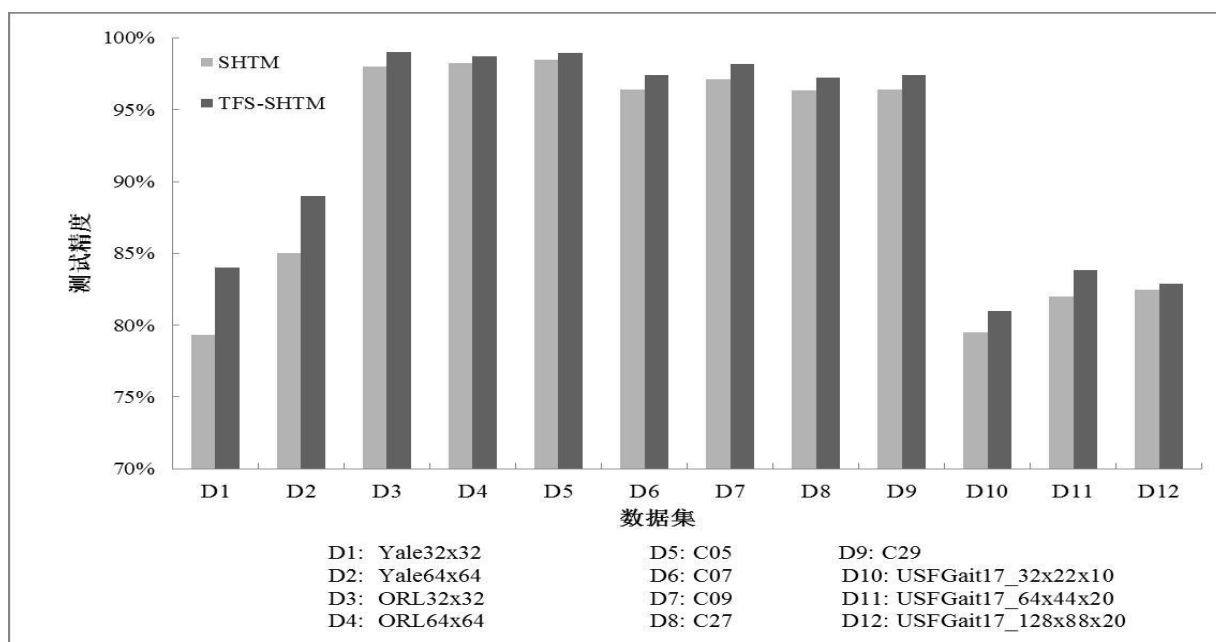


fMRI数据的可视化

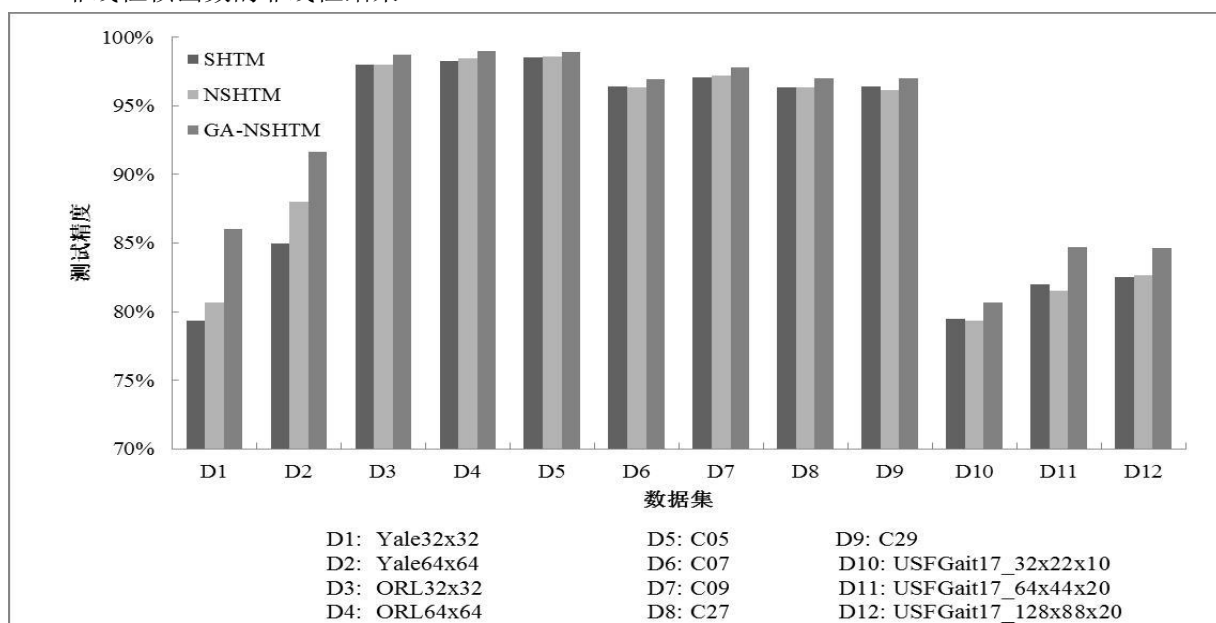
实验中使用的数据集

数据来源库	数据集	样本数	类别数	维度	数据样例
Yale-B	Yale32x32	165	15	32x32	
	Yale64x64	165	15	64x64	
ORL	ORL32x32	400	40	32x32	
	ORL64x64	400	40	64x64	
CMU PIE	C05	3332	68	64x64	
	C07	1629	68	64x64	
	C09	1632	68	64x64	
	C27	3329	68	64x64	
	C29	1632	68	64x64	
USF HumanID	USFGait17_32x22x10	731	71	32x22x10	
	USFGait17_64x44x20	731	71	64x44x20	
	USFGait17_128x88x20	731	71	128x88x20	

线性核的实验结果



非线性核函数的非线性结果



6.4 表

表直接显示数据，当需要数据的确切数值时，表比图表更可取。不过，表格的目标通常与数字类似：支持比较。

当以两个或多个维度显示数据时，表条目的布局和顺序会极大地影响读者进行适当比较和查看重要趋势的能力。

对于读者来说，比较排在行中的数字比列中的数字更容易。比较相近（最好是相邻）的数字也更容易。

同数字一样，表格本身也应易于理解，如果可能的话，不必参考文件的文本。这意味着一个表应该有一个好的标题，并且显示的项目应该在表中明确定义。

不要忘记单位和不确定度估计。

不同的期刊对表格有不同的格式要求。

一些例子：

学习 算法	数据集	测试 精度	训练时 间	学习 算法	数据集	测试 精度	训练时 间
SVM	Yale32x32	77.33	0.642	SVM	C09	97.40	584.664
STM		74.00	1.383	STM		96.23	655.519
SHTM		79.00	0.078	SHTM		97.45	49.128
SVM	Yale64x64	84.33	1.708	SVM	C27	96.69	348.773
STM		82.33	6.466	STM		95.10	653.308
SHTM		85.33	0.544	SHTM		96.72	68.924
SVM	ORL32x32	97.75	5.311	SVM	C29	96.62	298.991
STM		97.00	7.314	STM		94.75	631.321
SHTM		98.00	0.413	SHTM		96.64	90.223
SVM	ORL64x64	97.75	17.997	SVM	USFGait17_ 32x22x10	76.39	265.730
STM		96.50	34.299	STM		78.79	834.333
SHTM		98.50	3.208	SHTM		79.60	19.294
SVM	C05	98.59	2398.530	SVM	USFGait17_ 64x44x20	77.53	2896.670
STM		98.06	3129.298	STM		--	--
SHTM		98.76	203.475	SHTM		81.55	28.980
SVM	C07	96.47	324.912	SVM	USFGait17_ 128x88x20	77.53	8940.456
STM		95.44	648.103	STM		--	--
SHTM		96.74	34.158	SHTM		82.60	55.298

Datasets	Test Accuracy(%)		
	MPCA	OMPCA-s	OMPCA-m
	+SHTM	+SHTM	+SHTM
Yale 32x32	76.67	76.00	76.00
Yale 64x64	82.00	80.00	80.00
ORL 32x32	98.00	98.00	98.00
ORL 64x64	97.50	97.50	97.00
C05	98.94	98.82	98.88
C07	97.00	96.95	96.76
C09	97.75	97.65	97.65
C27	94.71	94.71	94.71
C29	95.59	95.59	95.59
USFGait17_32x22x10	67.38	66.15	66.26
USFGait17_64x44x20	74.59	72.98	72.91
USFGait17_128x88x20	77.70	73.95	74.34

与不同核函数的测试精度比较

数据集	平均测试精度(%): 均值±标准方差			
	DuSKRBF	RBF	Factor	K3rd
ADNI	0.75±0.18	0.49±0.23	0.51±0.21	0.55±0.14
ADHD	0.65±0.01	0.58±0.00	0.50±0.00	0.55±0.00

HIV	0.74±0.00	0.70±0.00	0.70±0.01	0.75±0.02
与不同核函数的训练(测试)时间比较				
数据集	平均学习时间(s): 训练(测试)			
	DuSKRBF	RBF	Factor	K3rd
ADNI	0.10(0.05)	2.22(1.09)	58.44(28.21)	25.18(12.15)
ADHD	2.20(1.09)	57.16(27.61)	1054.66(519.71)	635.19(315.23)
HIV	0.45(0.22)	16.12(7.81)	226.32(113.07)	190.21(94.32)

7. 作者身份 (Authorship)

科学论文中所报告的工作谁值得称赞? 这是科学著作权的基本问题, 因为与创造性写作中的著作权信用不同, 对科学论文来说, 最重要的是思想而不是文字。

7.1 定义

以下是适用于科学论文的作者定义:

一篇科学论文的作者是指那些对被作为新颖的词语或想法做出创造性贡献的人。

我们现在对作者身份有了一个定义, 并理解了与此相关的责任。基于这些前提, 以下是作者身份的三部分测试 (校验):

(1) 此人是否对工作做出了创造性的贡献?

请注意, 贡献包括撰写手稿和/或参与构思、设计、执行或解释作品。创造性的贡献是一种智力上的贡献, 它能增强作品的新颖性。

(2) 在提交出版之前, 此人是否审阅并批准了最终的手稿?

(3) 这个人是否接受作者身份带来的责任, 包括愿意和有能力回答批评?

一些对论文很重要但没有做出创造性贡献的作品的例子包括:

使用标准方法准备材料或操作设备, 即使此类工作是广泛的;

应用常规统计检验或分析而不作解释;

对稿件的日常审查、校对或编辑;

监督参与工作的人员, 批准他们的项目, 或获得资源。

7.2 No Guests or Ghosts

在列出手稿的作者时, 有两种方法是错误的:

撇开属于清单的人 (实际作者, Ghost)

包括不属于名单的人 (客串作者, Guest)。

由于不同的原因，这两种错误在科学出版中都是相当常见的，而且这两种错误都可能是严重的问题，后果也不同。

通常，这样的错误是无意的，通常是由于不完全了解对作者身份的要求造成的。

不过，有时候，这个错误并不是那么无辜的，而是道德上的严重违背。

7.3 别忘了致谢

大多数作者会在最后一刻为论文考虑一个致谢部分。“别忘了提及我们的资金资助机构，”合著者之一在一份迟交的草稿上匆忙记下。

然而，致谢所有对作品有贡献，但其贡献没有上升到作者级别的人是极其重要的。

这里列出了技术人员、主管或同事，他们的工作很重要，但不是论文新颖方面的一部分。

如果你考虑过将某人列入作者名单的可能性，但最终没有这么做，那么此人很可能属于“致谢”部分，并说明他们的贡献。

7.4 作者顺序

第一作者通常被认为是获得最多荣誉和责任的人。

然后根据作者对作品贡献的降序排序，贡献越大，排序越靠前。

但是不同的社区有不同的文化，这种作者排序系统并不普遍。

另外两种确定作者顺序的系统已经很常见：

(1) 第一种方法是通过始终按字母顺序列出作者，将作者顺序与贡献级别断开。数学期刊的文化是按字母顺序列出作者，这种做法几乎是普遍遵循的。许多数学论文只有一个或几个作者，这一事实可能使这种做法更容易采用。

(2) 当出版涉及博士生或博士后的工作时，另一种系统相当常见。在这里，作品通常代表一个学生的论文项目，然后分配给他第一作者位置。那个学生的导师被指派为最后一位作者。

8. 双重出版

同行评议期刊几乎总是有一个限制，对双重出版物提交的手稿，是基本上相同的一个已经发表了另一个同行评议期刊。一个相关的概念（这也是禁止的）是双重提交，其中相同或基本相同的手稿正在考虑由两个同行评审期刊同时出版。

8.1 旧的东西，新的东西

除其他标准外，手稿必须包含新颖的内容，以使其可以在同行评审的科学期刊上发表。但是，并非论文中讨论的所有内容都必须新颖。通常，论文会先讨论先前（已发表）的结果，然后再讨论新内容。明确区分先前的工作和新的结果是作者的责任。可以通过直接语言（“先前的工作已经显示出来……”；“在这项工作中，我们进行了测量……”）来明确地做到这一点，也可以通过引用来隐含地做到这一点。

8.2 会议记录的作用

光电子工程协会（SPIE，国际光学工程学会）对于将会议论文的内容提交给其同行评审期刊之一具有相当宽容的政策。原因很简单：SPIE 认识到会议及其程序在科学知识增长中的重要和独特作用，是对同行评审期刊重要作用的补充。

我们的理念是会议和期刊应该一起工作，而不是竞争。会议记录提供了会议记录，是科学或工程领域迅速发展的一個简要说明。经同行评审的期刊对完成的工作（或至少在更大的工作中至少是一个里程碑）进行了异步的观察，并仔细地介绍了它们，为科学界提供持久的

价值。

9. 抄袭

剽窃（抄袭，**Plagiarism**）通常被定义为将他人的想法、图像或文字表现为自己的想法、图像或文字。这是智力盗窃。尽管这个定义看起来很明显，但根据我的经验，定义和识别剽窃的实践要复杂得多。通常，剽窃更多的是智力懒惰的结果，而不是智力上的不诚实。

9.1 抄袭他人的想法

每项新作品都以过去的作品为基础，这是科学的基石原则，因此，不仅允许而且鼓励利用他人的思想。道德失误（小错）源于将这些想法错误地陈述为自己的想法。这种错误陈述可以是明确的（“我们第一次来这里……”），但通常是隐性的。通过不加引述地提出想法，设计，模型，过程或结果，就可以清楚地看出这些想法是原始的。

9.2 抄袭他人的图像

图像是科学交流的重要组成部分，图像或其他图片的产生通常是一种创造性的行为。因此，使用他人的图形不仅需要引用其原始出版物，还需要图形的作者（可能还有出版者）的许可。

9.3 抄袭别人的话

文本复制行为的严重性可能会发生很大的变化，从整份论文的整体抄袭到一些句子的措词。虽然绝对不允许复制文本（不带引号），但问题的严重程度取决于几个重要因素：

是否引用了复制文本的来源？与粗心或不良的写作习惯相反，缺乏引用被认为是欺骗意图的证据。

复制了几句？复制量越大，冒犯性越大。

是否有释义（**paraphrasing**），或仅仅是掩饰（**disguise**）的企图？可以通过适当的措辞避免抄袭。但是，对复制的句子中的一两个词稍作改动，就与用自己的单词重写不同。请注意，改写段落仍然需要引用原文。

是否有任何抄袭的文章被认为是论文的新颖之处？仅限于背景材料或方法部分的复制仍然是窃取，但不如复制结果或解释那样糟糕。

9.4 重复发表或自我剽窃

“自窃”一词是矛盾的修辞法（**oxymoron**）：您不能从自己身上窃取。尽管如此，该术语仍经常被用来描述一个严重的问题：将以前发表的作品误称为新作品。作者有时会使用这种没有适当引用的重复出版物来增加出版物数量，希望编辑和审稿人不会注意到最新论文中缺乏新颖性。这对期刊及其读者是有害的，他们浪费时间去阅读和阅读旧著作，认为有一些新东西需要学习。因此，重复出版是严重的道德违背。

如果您复制自己的文本或图形，请引用该文本或图形（尽管您无需添加引号）。如果您的新作品是旧作品的延续，请引用它。确保读者可以轻松区分新内容和旧内容。

10. 选择要发表的期刊

在科学家或工程师考虑发表同行评议论文时一个重要的问题是要发表到哪个期刊。这个问题从一开始就在研究者的脑海中。

10.1 选择对的/合适的期刊

哪些因素会影响你决定你的研究工作要发表的地方呢？在以前，期刊选择涉及到相关性、该期刊的接收率、发行量、声望以及出版时间。但是随着越来越多期刊能在网上找到，搜索引擎使得查找和访问文章变得更加容易，其中的一些因素在如今已经不那么重要了。

10.2 避免错的/不合适的期刊

为了避免无德期刊，这里有一个问题清单，当你在不熟悉的期刊上提交论文前，你可以自己尝试回答这些问题：

你或你的同事知道这本期刊吗？你以前读过这个期刊的文章吗？在期刊上找到最新的论文容易吗？

你能很容易地识别和联系这些出版商吗？期刊网站上是否清楚地显示出版社的名称？你能通过电脑、电子邮件或邮寄的方式联系到出版商吗？

期刊是否清楚它所采用的同行评议的类型？

在你使用的服务中是否为文章编制了索引？

收费的标准明确吗？期刊网站是否解释了这些费用的用途以及何时收取？

你认识编委会吗？你听说过编委会的成员吗？编委会成员是否在自己的网站上提及该期刊？（有时未经过本人的允许，有些人会被列为编委会的成员。）

11. 编辑看什么

要想在科学期刊上发表，论文必须满足四个重要标准：

- （1）论文的内容必须符合期刊的范围。
- （2）论文的质量（研究的方法和执行以及写作）必须足够高。
- （3）它必须提出新颖的结果（综述等除外）。
- （4）结果必须足够重要，值得阅读（从而值得发表）。

11.1 范围（Scope）

稿件被拒的最简单方法就是投错期刊。如果稿件的主题不符合期刊的范围，再好的稿件也会被拒绝。因此，你应该仔细研究你想投稿的任何期刊的范围，并确保有匹配的范围。

11.2 质量（Quality）

与期刊出版物相关的质量有两个方面：一是所报道工作的质量，二是报道的质量（即书面稿件）。工作质量基本上是对所涉科学的判断，包括在规划和执行实验、分析所产生的数据以及将这些结果纳入科学领域的大框架方面所采取的谨慎态度。

11.3 新颖性（Novelty）

除了综述性论文和教程外，稿件必须包含新的内容，才值得在科学期刊上发表。

因此，一篇期刊论文必须为该知识体系增添一些新内容（新理论、新设计、新模型、新方法、新数据或新分析）。因此，有效的文献检索和全面的引文是确定所提交作品的新颖性的必要条件。

一个好的标准（good rule of thumb）是，至少 50% 的成果必须是新的。

11.4 重要性（Significance）

最后的发表要求也许是最模糊的（nebulous）：作品必须具有足够的意义。意义的判断应该基于读者的观点：有多少人会阅读这篇论文，并将所传达的知识用于实践。

编辑和审稿人很难前瞻性地判断投稿的意义。一般来说，编辑和审稿人会采取两步走的

方式来进行这样的评价。工作所解决的问题有多重要，以及这项工作比以前的文献有多大的进步？

11.5 结论

期刊编辑总是在每一篇投给他们期刊的稿件中寻找四点：范围、质量、新颖性和意义。在提交稿件发表之前，可以试着用自己用这四项来评估一下。

12. 推荐信（cover letter）

12.1 推荐信的目的

当总编辑审阅稿件时，他的第一个决定是他认为稿件经过同行评审是否会很有成效，还是应该不经审查就予以拒绝。推荐信为他提供了进行第一次重要评估（或至少应该评估）所需的信息。

为什么不经审查就可以拒绝稿件？有三个基本原因。

（1）首先，该论文可能不属于我们期刊的范围。（如果可能，主编会尝试推荐一种更合适的期刊，并鼓励作者尝试在那里）。

（2）拒绝审核的另一个原因是稿件的英文不好。如果他由于这个原因而拒绝，他鼓励作者由以英语为母语的人（或可选地，由商业编辑服务）编辑手稿，然后重新提交。

（3）最后，如果很明显论文不是新颖的或不重要的，编辑可以拒绝发表论文。

12.2 推荐信的结构（内容）

（1）稿件信息：提交的稿件标题和文章类型（信件，常规论文，特别论文，评论，教程，交流等）。如果要提交特别部分，请提及特别部分的名称。

（2）解决的问题：哪些问题促成了这项工作？这项工作旨在填补什么空白？这项工作的大背景是什么？

（3）作品的新颖性：这里有什么新内容，以前没有发表过？“据我们所知，这是第一份显示……的报告。”

（4）作品的意义：本文的什么新内容是以前没有研究讨论过的？对这个领域的潜在影响是什么？

（5）适合该期刊：为什么这项工作属于该期刊并吸引读者？该手稿的出版将如何使该期刊受益？（请熟悉该期刊的范围。）请说明这篇论文是建立在之前发表的论文的基础上，还是与之前发表的论文直接相关。

（6）双重出版：“该手稿先前未曾出版过，目前尚未在印刷中，正在审查中，或正在考虑由另一种期刊出版。”

（7）作者批准：“所有作者均已阅读并批准了所提交的手稿，并同意将其提交至本期刊。”

最后，以标准信件书写格式结尾，注明相应作者的姓名和联系信息。

13. 编辑审查过程

同行评议（peer review）是大多数科学期刊出版过程中的一个关键环节。然而，对许多作者来说，编辑审查的过程可能看起来有些吓人（intimidating），甚至有点神秘。

同行评议被定义为“由通常不属于编辑人员的专家对提交给期刊的稿件进行批判性评估”。

它支持科学进程，为作者提供对其作品的建设性批评，并过滤掉不太有价值的作品，从而为已发表的科学作品提供编辑和同行的“认可印章”。

仅仅是因为需要同行评议，就促使作者在提交的手稿中改进科学研究及其呈现方式。

13.1 同行评议的目标

两个直接的目标：帮助编辑决定发表哪些手稿，拒绝哪些（过滤）；为作者提供如何改进论文的建议（批评）。

此外，在同行评议期刊上发表的“认可印记”可以帮助作者的职业生涯。

编辑同行评议过程中伴随的过滤和批评有助于将最好的论文有效地送到最感兴趣的读者手中。

为了使同行评审过程实现其目标，评审必须具有良好的质量。

任何发表过相当数量论文的人都知道有些评论的质量要比其他的高很多。

一篇好的评论教给作者写作和科学知识，不仅能写出一篇更好的论文，而且还能使作者以后的每一篇论文都写得更好。这也使编辑的工作大大简化。

13.2 优秀评议的特点

事先说清楚，编辑和审稿人在撰写评论时都不需要使用正式的检查表。

有时，为了帮助编辑和审稿人确保科学论文的最重要方面得到考虑，会给出一个参考准则。

对一个作家来说，像一个读者一样思考总是一个好建议，而第一个读者就是编辑和审稿人。

在阅读并批判性地评价一篇稿件之后，审稿人现在必须把这种评价转达给期刊编辑。在任何情况下，都应该使用尊重和建设性的语气。审查的形式并不重要，但每次审查都应包含某些重要信息。

格式：

第一段应包含以下三个要点：

- （1）提供一篇简短的（1-2 句话）论文摘要；
- （2）解释这篇文章的新颖之处（1-2 句话），包括作者的观点和你的评价；
- （3）解释为什么这项工作很重要（1-2 句话）。

第二段应概述所报告研究的质量。如果在从方法到数据到分析再到结论的逻辑过程中存在任何重大缺陷，请在这里提出，并说明可以采取什么措施来修复这些缺陷。在这一段中，把重点放在大问题上（如果有的话）。

审查的第三部分，也是最后一部分，应该列出作者应该解决的具体问题。这些点可以小或大，从图片格式到论文结构。请记住，稿件编辑工作将由期刊工作人员在验收后完成，所以不要担心语言或格式的问题，除非它们妨碍了你正确理解和审阅稿件的能力。

13.3 同行评议的过程

期刊应该有一个完善的同行评审程序。以下是对期刊使用的稿件审查过程的一步一步的描述。

- （1）作者在网上提交他们的手稿，以及一封推荐信（cover letter）和其他各种信息。在这一步中，作者为他们的手稿选择一个当前开放的专区或常规的论文或信件类别。
- （2）稿件经过期刊工作人员的质量控制检查。如果提交有问题，期刊的工作人员会与相应的作者一起解决这些问题。
- （3）稿件通过相似性和抄袭检查处理，将提交的稿件与先前发表论文的大型数据库进行比较。如果提交的材料中有足够的内容与先前发表的论文中的内容完全相同，我们会联系作者

进行解释，如果发现了严重的问题，可能会拒绝原稿并实施制裁（**sanction**）。

（4）根据投稿过程中的类别选择，稿件将交给特约专区的客座编辑或与常规论文类别相关的高级编辑（**Senior Editor, SE**）。编辑通过阅读投稿信、标题和摘要，并浏览论文进行第一次编辑评论。这位编辑检查论文的范围是否与期刊的范围相匹配，以及文章写作是否足够好，以便进行有效的评论。如果没有，编辑可以决定不经审查而拒绝稿件。

（5）对于常规的提交，高级编辑决定由一名具有适当专业知识的副编辑（**Associate Editor, AE**）来处理提交。副编辑不一定是该主题的专家，但他们会非常熟悉并能够找到评审人并解释他们的评论。对于专区的提交，客座编辑们将决定哪位客座编辑将担任本次提交的副编辑的角色。

（6）副编辑负责期刊的大部分编辑工作。他们首先对论文进行第二次编辑审查，检查范围、新颖性、重要性和质量。他们可以快速浏览论文，也可以详细阅读。副编辑必须决定论文是否有机会被接受发表，并决定因此是否值得送审。

（7）如果副主编不决定不经审查就拒绝原稿，他们将寻找并指派合格的审稿人。接受一份手稿出版至少需要两次评审，但一些副主编可能会选择寻求三次评审。通常，审稿人被选择具有互补的技能（实验、理论、数学等），以便手稿中的所有主题都能有专家分析。作者可以在提交时提供一份推荐评审人的名单，但是否使用该名单中的任何人由副主编决定。在这个过程中，找到合格的评审员通常是最困难和最有可能产生问题的一步。有时在两名评审员接受任务之前，必须询问 10-20 名候选人。

（8）当评审被反馈后，副主编会评估这些评审并做出决定（通常是要求作者修改或拒绝）。尽管审稿人可能会提供接受/拒绝的建议，但副主编会根据他们对原稿的阅读和评审的实质内容做出最终决定。

（9）如果作者修改了手稿，它将被送回到同一个副主编。副主编会查看修改后的手稿以及作者对审稿人评论的逐点响应，然后决定将手稿送出重审，或者此时做出接受/拒绝的决定。根据修订的程度，可以进行多轮重新审查。一般来说，手稿将被发回给相同的审稿人，但如果原始审稿人没有空，或者如果大量增加的材料需要一位具有额外专业知识的审稿人，则可能会选择新的审稿人。

（10）最后，在稿件决定作出后，建议的决定将送交总编辑批准。总编辑对整个编辑过程进行最后的质量检查，可能会提出修改或改进的建议。很少有主编会以任何方式改变副主编的决定。

（11）如果稿件被接受，作者将收到关于手稿及其图表如何最终提交的指示。接受稿件后将不得再更改稿件的内容。

（12）最后提交的手稿经过文案编辑和专业写作步骤。这些重要且通常不为人所知的步骤可能会对论文的专业水平产生重大影响，纠正打字错误和语法错误，改善论文的阐述和呈现方式，并确保图形具有足够的质量。

（13）书页校样被发送给相应的作者以供批准，并可能提供缺失的信息。作者应及时回复这些证明。

（14）论文定稿后立即在网上发表，并在本季度末刊登在期刊的印刷版上。

一个同行评议的例子（[附件五](#)）。

13.4 责任

作者

（1）确保作品是原创的，之前没有出版或提交其他地方出版。正确引用你自己之前和重叠的作品。

- (2) 选择合适的作者列表，并由所有作者完全批准提交的内容。
- (3) 选择最合适的期刊，提交最好的手稿。千万不要故意提交一份拙劣的手稿，希望编辑和审稿人能帮你修改。
- (4) 花时间了解所选期刊的提交要求，并遵守这些要求。
- (5) 确定所有的资金来源，并让编辑意识到任何潜在的利益冲突。

编辑

- (1) 提供一个透明的编辑审核流程，只有在特殊情况下才偏离该流程。
- (2) 在出版过程中公平和尊重各方。
- (3) 当你处理一个你有利益冲突的稿件时，回避并让一个没有冲突的编辑来处理你的稿件并做出决定。
- (4) 确保提交文件的所有细节都是保密的。
- (5) 刻苦工作以便及时做出决定。
- (6) 选择有可能提供公正、公正、高质量和及时评论的审稿人。
- (7) 使出版过程中的各方遵守最高道德标准。

同行评审员

- (1) 披露任何可能会使你对原稿的意见产生偏见的利益冲突（由竞争、合作、财务或其他关系引起的）。如果你在利益冲突的情况下被选中进行审查，请尽你所能提供一个公正的评价。
- (2) 快速返回评论。如果您由于任何原因无法及时返回质量审核，请尽快通知编辑。
- (3) 提供一个建设性的，专业的评论，它不应该是针对个人的。
- (4) 提供详细的评论，用证据支持所有观点；你的目标应该是帮助作者改进他们的论文，即使你建议拒绝。
- (5) 对从你的审查中获得的信息保密（**confidential**）。在手稿出版之前，不要透露或使用从审稿中获得的知识。

13.5 对同行评议过程的一些批判性思考

同行评议的过程也有其批评者，其中一些人的声音相当响亮。以下是针对同行评审过程的一些主要批评：

它通过拒绝不一致的（**non-conforming**）或有争议的（**controversial**）观点来扼杀（**stifle**）创新，并通过拒绝无效的结果来扭曲记录。

它是不可靠的，经常找不到工作中的主要缺陷，包括欺诈和剽窃。

它既不一致也不客观，而且常常在一些方面有偏见。

出版是昂贵的并且有延迟。

几乎没有证据表明它是有效的，更不用说最好的方法了。

大多数被拒绝的文章最终发表在另一个同行评议的期刊上。

14. 综述（review）

14.1 什么是综述

对第三项要求——新颖性而言，综述论文这类重要的论文类别是一个例外。综述论文，顾名思义，是对某一特定主题之前发表的作品进行批判性评价。综述往往很受读者欢迎，因为它们在一个小空间里包含了大量信息，给读者投入的时间带来了巨大的回报。它们是一份礼物。

每次综述的主要目标应该是围绕选定的主题对过去的工作进行组织和综合，以便加速将最近的知识积累和吸收到现有的知识体系中。每一篇综述论文都应该有一个故事，一个主题和一个观点。它应该以创意为导向，而不是以文学或作者为中心。

下面是一些最好的综述论文中常见的主题：

争论：两个或两个以上的阵营对一种现象提出相互竞争的理论或解释，每个阵营都有依据。

关于主要新工具、工艺、方法或理论的进展。

导致重大发现或重要概念诞生的历史发展过程，及其对今天和未来的影响。

比较不同的方法对测量/设计/装配（**fabrication**）/建模具体的事物的重要性，及其优缺点。

跨学科或针对不同应用使用特定工具/过程/方法。

从更广泛的角度对某一话题最近的进展获得的新见解，或是对一个关键的新问题或之前未被注意的问题的认识。

号召行动：为什么社区应该在某一主题上投入大量资源。

14.2 综述的结构

综述论文的引言类似于研究文章的引言。它以描述背景主题和为什么这个主题是重要的作为开始。它说明了对这一专题的知识方面的空白，而将被综述的最近的工作填补了这一空白。然后概述当前综述的主题（争议、进展、历史发展等），以及如何与该主题相适应，以及知识差距（缺陷）。重要的是，引言要清楚地定义综述的范围，这样读者才能知道哪些内容是需要考虑的，哪些是不需要考虑的。

中间部分：一篇综述论文的中间部分的结构是根据所讲述的故事设计的，因此很大程度上取决于所选择的主题。一个好的作者会让故事引导评审的流程，始终牢记组织和综合的目标。

结语部分：与引言一样，综述论文的结语部分与研究论文的结语部分相似。结论一般化，寻找可以传授的更大的教训。在非常简短地总结了综述和它的主要信息之后，应该强调审查工作的含义，并指出我们目前的知识中仍然存在的差距（缺陷）。一般来说，读者之后会期望关于未来被需要的工作的描述和未来的有关问题的回答。以一些猜测作为结尾是好的，只要它被贴上这样的标签。

14.3 怎样才算一篇好的综述

综述应该是带有批判性的但公平的，而不是仅仅接受之前发表的论文中的结论。但不要涉及个人：在批判时就事论事，批判作品而不是作者。

通常，作者会将他们自己的工作作为综述的一部分。毕竟，因为他们对该领域有贡献，所以作者通常是被综述领域的专家。为了减轻这种感知到的利益冲突，在将作者自己的作品融入该领域的整体文献时，必须实现一种困难而谨慎的平衡。

15. 科学发表的伦理问题

正如本课程中多次提到的，科学论文写作的主要精神是以读者为中心，而不是以作者为中心。

15.1 科学出版的首要伦理

要使一个结果具有科学性，并对科学知识体系作出贡献，就必须对其进行充分描述，以便该论文的结论能够被其他人验证。

研究人员称之为科学出版的首要伦理。

这一伦理允许新的科学知识增加现有的知识，并促进科学的发展。

当商业或竞争利益介入时，作者可能会受到压力而不能在论文中提供足够的细节。公司可能希望将某些想法作为商业机密。作者可能想要隐藏缺陷，以增加发表的机会，并最大化其重要性。作者也可能希望对某些技术保密，以便在产生新结果方面领先于竞争的研究小组。秘密可能是令人渴望的，甚至是必要的，但它们不是科学的一部分。

简而言之，如果其他利益要求证明一篇论文结论的必要细节不能公开，那么这篇论文就不应该发表在同行评议的期刊上。想要隐藏必要细节的作者不应该提交这样的作品来发表。

15.2 出版前作者责任

在向同行评审期刊投稿之前，作者的主要职责如下：

开展研究导致在一个道德的方式出版。

写你的论文与开放和诚实，保持科学出版物的主要伦理。

写作时引用来避免剽窃通过马虎的（sloppy）引用实践。

确保工作是原始和以前没有发表或投稿。恰当地引用自己之前和重叠的工作。

选择适当的作者列表，提交由所有作者的全部批准。

选择最合适的期刊和提交最好的手稿。永远不要故意提交一份差的稿件，希望编辑和审稿人能帮你修改它。

花时间理解选择期刊的提交需求和符合这些要求。

识别所有资金来源并通知编辑任何潜在的利益冲突。

15.3 同行评审过程中的作者责任

在此过程中，作者的主要职责如下：

尊重编辑出版人员在整个出版过程。

不采取关键评论个人（这可以很难遵循的建议），和从来没有回应记者的评论而生气或沮丧。把对你工作的批评理解为你自己的批评是人的天性，但这很少是准确的回应，也从来都不是恰当的回答。如果你在写评论回复的时候发现自己的体温上升了，那就把它写下来，以后再做。

几乎在所有情况下，修订以回应评论将使论文更好。尽管你可能会情绪反应和修改所需（entail）的额外工作，感激这个机会改善你的论文基于专家的评估。

回复日记手稿修改请求通过提供一个逐点详述的应对每一项由评论家和编辑。你不必接受审稿人提出的每一个修改请求，但是如果你不同意某一点，解释为什么（如果合适的话，用证据）。如果你根据审稿人的观点对稿件做了修改，请准确描述所做的修改。

提交修订后的手稿给期刊之前，确保每个作者已经批准所有的变化。

在罕见的情况下，材料添加到修改手稿可能需要添加一个新的合作者。如果是这样，请在回复中仔细解释为什么要添加新作者。

记住在同行评审过程中发现的材料在你的手稿不能提交给另一个杂志上考虑。如果你的手稿被拒绝，你可以自由地将手稿提交到其他地方。然而，非常明智的做法是，认真对待任何与拒绝有关的评论或批评，并在再次尝试之前相应地改进你的手稿。

15.4 出版后作者责任

作者的责任并不随着出版而结束。以下是作者出版后的主要职责：

作者负责应对雷声大雨点小的批评他们的工作后出版。如有必要，发布后发现的错误

应通过勘误表或后续发布更正。

准备共享数据中发现你的论文（或你的结果依赖）其他研究人员。一旦出版，你必须考虑这些开源数据而不是让这些数据保持专有的（**proprietary**）。

因为你可能需要分享，所以本文依赖的所有数据应精心组织和存档，保存时间越长越好（至少三年是一个很好的目标）。

16. 编辑的伦理准则

本课程有几章涉及作者的道德责任：如何正确引用他人作品，如何确定作者名单上的作者，如何避免抄袭和重复出版。在这里，我将更详细地阐述编辑的道德责任。

16.1 编辑的责任

尽管有许多方式来总结在同行评议科学期刊工作的编辑的道德义务，以下是 Chris A. Mack 总结的 7 项要点：

- （1）为编辑审查提供一个透明的（**transparent**）过程，并避免只在异常机制下的流程。
- （2）公平地和尊重地应对出版流程的各方。
- （3）当处理有利益冲突的手稿时撤换自己（**Recuse yourself**）——让一个没利益冲突的编辑处理投稿并作出决定。
- （4）对投稿的所有细节保密。
- （5）勤奋工作（**Work assiduously**）以及时作出决定。
- （6）选择公平、无偏见、高质量且及时的审稿人。
- （7）将出版流程的各方保持在最高的道德标准。

17. （对学生的）一些建议

认真阅读你所研究领域的论文，并用 1-2 句话进行评论。

认真推敲论文的题目、摘要、综述和结论。

对于论文的创新点，合理设计实验，从不同的方面验证方法的有效性。

对于实验部分，一定要和最新和重要的参考文献做比对性工作，尊重别人的劳动成果。

天道酬勤，希望同学们在做研究的过程中专注、持续、认真。

参考资料

- [1] 杨晓伟副院长的课件
- [2] 学术论文（总）.docx
- [3] 论文写作(1).pdf

附件一：介绍的修改样例

原始的介绍：

Tensor analysis as a powerful tool for multiway systems has been applied to variable fields, such as computer vision [2][3], machine learning [4][5], neuroscience[6], signal processing [7], data mining [8][9] and so on.

Nowadays many multiway systems have suffered from missing data, and tensor completion with missing data has drawn more and more interest [17]-[27]. As we can see, most researches on this issue share a common assumption that the original tensor is of low-rank.

Some researches on tensor completion are based on tensor factorization, which plays an important role in multiway data analysis. The most important tensor factorization methods are CANDECOMP/PARAFAC (CP) factorization[28] and Tucker factorization[14][15]. To address the missing value problem, many classical tensor factorization methods have been extended to incomplete tensor cases. Based on CP decomposition, Acar et al. [17] proposed CP-WOPT and solved the model by the gradient descent algorithm. Based on Tucker decomposition, Geng et al. [18] proposed M2SA for extending N-mode dimensionality reduction [19] to the incomplete tensor cases, and used an iterative process to get the final solution. Filipović et al. [20] proposed WTucker and the model is solved by the nonlinear conjugate gradient (NCG) method. Recently, based on Cauchy distribution, Wu et al. [22] presented a robust tensor decomposition framework for recovering a low-rank tensor, and developed two algorithms CTD and CCD based on Tucker and CP respectively. All of these methods are based on tensor factorization, and usually pick the exact tensor rank from some range. With the chosen small enough, the recovered tensor will be of low-rank. But the real rank of the tensor is not easy to be estimated.

Another way to recover low-rank tensor makes use of traditional matrix completion methods. For example, nuclear norm is originally a concept from matrix analysis. It's been proven the tightest convex surrogates of the matrix rank norm[29]. Thus, the minimization of the rank of a matrix can be alternatively transformed into the minimization of its nuclear norm, which is much easier dealt with compared to the rank norm. As far as we can see, Liu [23] firstly extends the matrix nuclear norm to the tensor cases, and then uses the tensor nuclear norm regularization to force the tensor rank shrink. This method unfolds the tensor in each mode, and penalizes the nuclear norm of the unfolding matrices. Finally the optimal issue is solved by ADMM. Under the framework of [23], a few models are developed in [24][25][26][27]. Among the models, by use of the inequation that the rank of a matrix is smaller than its factor matrix, TNCP [27] successfully introduces CP factorization into the framework.

Tucker factorization method is usually solved by the alternating least squares (ALS) algorithm called HOOI[30], which is an iterative process. As the non-truncated version of Tucker, HOSVD[16] follows HOOI but without iterations, and thus takes much less time compared to Tucker. At present, by simply restricting the factor matrices of HOSVD to their leading columns, the truncated HOSVD (T-HOSVD) obtains a good approximation of Tucker factorization. In order to obtain the higher accuracy, Vannieuwenhoven [1] proposed a new truncation method for HOSVD called sequentially truncated HOSVD (ST-HOSVD), in which the factor matrices are obtained sequentially in a greedy way. Due to the greedy strategy, it's quite fast with remarkable accuracies. Compared with the other tensor factorization methods like Tucker, CP and T-HOSVD, another advantage of this novel method is that one can conveniently gives its error analysis.

Considering that most tensor factorization methods need to specify the tensor rank as input and it is very difficult to give the best tensor ranks, in this paper, we firstly present an adaptive method based on ST-HOSVD to deal with tensor factorization. As for the tensor rank, we adaptively restrict the tensor rank by controlling an parameter which we call the error tolerance, instead of picking the rank directly.

With this adaptive tensor factorization method, we generate a series of low-rank approximation of a tensor, and take the average as an approximation of it. Finally, for incomplete tensors, we present an iterative algorithm based on the adaptive tensor factorization method to fulfill it.

Thanks to ST-HOSVD, our method is extremely fast in many cases. In the experimental result, our method also obtains good accuracies, and we think this is because of the average operator we use to get the approximation of tensors. Note that the average operator is one of the most popular ways to reduce global error.

修改后的介绍:

Tensor analysis as a powerful tool for multiway systems has been successfully applied to computer vision [1-4], audio processing [5], machine learning [6-11], neuroscience [12], geophysics [13], signal processing [14-15], data mining [16-20] and so on. Tensor factorization plays an important role in multiway data analysis. For complete tensor, the most important tensor factorization methods include CANDECOMP/ PARAFAC (CP) factorization [21-25], Tucker factorization [26-29] and higher order singular value decomposition (HOSVD) [30-33]. Mathematically, HOSVD is the special version of Tucker factorization. Without dimensional reduction, Tucker factorization is reduced to HOSVD. At present, Tucker factorization is usually solved by the alternating least squares (ALS) algorithm called HOOI [29]. By simply restricting the factor matrices of HOSVD to their leading columns, the truncated HOSVD (T-HOSVD) obtains a good approximation of Tucker factorization. Recently, in order to obtain better performance, sequentially truncated HOSVD (ST-HOSVD) [34] is proposed for approximating Tucker factorization, in which the factor matrices are obtained sequentially in a greedy way. Due to the greedy strategy, ST-HOSVD is much faster than HOOI with remarkable accuracy. In addition, compared with Tucker, CP and T-HOSVD, another advantage of ST-HOSVD is that one can conveniently give the approximation error analysis.

In real world applications, missing values may occur in data sets for a number of reasons: glitches and malfunctions of one or more sensors, irregular measurement intervals between samples, or different sampling frequencies for the various sensors. In the past several years, tensor analysis with missing data has drawn more and more interests in multiway systems. As we can see, most researches on this issue share a common assumption that the original tensor is of low-rank. In order to capture the low-rank structure of the incomplete tensor, many classical tensor factorization methods have been extended to incomplete tensor cases. Based on the Expectation Maximization (EM) approach [35], Bro [36], Walczak [37], Tomasi [38], Acar [39], and Karlsson [40] extended the CP factorization to missing values. The main difference between them is that [36], [37], [38], [39], and [40] used alternating least squares (ALS), the Levenberg–Marquadt method [41], the gradient descent algorithm and parallel algorithm to update the factor matrices, respectively. Recently, Shao [42] presented a CP factorization based sparse model for clustering multi-source incomplete data. Yokota [43] proposed an effective smooth CP factorization based on the smoothness constraints and low rank approximations for the case of the missing ratio being extremely high and Zhao [44] proposed a generative model based on Bayesian inference and CP factorization to deal with missing data and outliers. Motivated by the idea in [36-39], Geng [45] proposed a multilinear subspace analysis (MSA) method based on Tucker factorization for the incomplete tensor data and applied it to face recognition and facial age estimation.

Filipović [46] proposed the weighted Tucker factorization for tensor completion. Recently, based on Cauchy distribution, Wu [47] presented a robust tensor decomposition framework for recovering a low-rank tensor, and developed two algorithms Cauchy Tucker Decomposition (CTD)

and Cauchy CP Decomposition (CCD) based on Tucker and CP, respectively. At present, except for methods which use adaptive strategies to choose CP rank [43][44], other methods usually pick the exact tensor rank from some range. With the chosen small enough, the recovered tensor will be of low-rank. Unfortunately, estimating the ranks usually needs costly operations.

Except for tensor factorization based methods mentioned above, tensor nuclear norm is also used to recover the low-rank structure of incomplete tensors. The previous studies show that nuclear norm is the tightest convex surrogates of the matrix rank norm [48]. Based on this fact, Liu [49] extended the matrix nuclear norm to the tensor cases and used the tensor nuclear norm regularization to force the tensor rank shrink. Following the framework in [49], Gandy [50] dealt with the tensor completion problem with noise, and Liu [51] proposed a more general definition of tensor nuclear norm and applied it to low-rank tensor completion problems. The main drawback of [51] is that the algorithm involves many SVDs of very large matrices, and thus suffers from very high computational cost. To overcome this drawback, Liu [52] introduced CP factorization into this framework by using the inequality that the rank of a matrix is smaller than that of its factor matrix.

In this paper, an ST-HOSVD based algorithm is proposed to deal with tensor completion. Considering it not convenient to specify the ranks, in this study, we firstly present an adaptive method based on ST-HOSVD to deal with complete tensor factorization. As for the tensor rank, we adaptively restrict it by controlling an error tolerance parameter instead of picking the rank directly. With this adaptive tensor factorization method, we generate a series of low-rank approximations of tensor, and take their averages as its low-rank approximation. Finally, for incomplete tensors, we present an iterative algorithm based on the adaptive ST-HOSVD and the average operator to fulfill it. The experimental results on images and videos show that the proposed method has higher computational speed and obtains better accuracy compared with the other incomplete tensor algorithms.

翻译版本

原介绍:

张量分析作为多路系统的有力工具,已被应用于各种领域,如计算机视觉[2][3]、机器学习[4][5]、神经科学[6]、信号处理[7]、数据挖掘[8][9]等。

目前许多多路系统都存在数据缺失问题,而利用缺失数据进行张量补全已引起人们越来越多的关注。我们可以看到,关于这个问题的大多数研究都有一个共同的假设,即原始张量是低秩的。

张量补全的一些研究是基于张量分解的,它在多路数据分析中起着重要的作用。最重要的张量因子分解方法是 CANDECOMP/PARAFAC (CP)分解[28]和 Tucker 分解[14][15]。为了解决缺失值问题,许多经典的张量因子分解方法被推广到不完全张量情况下。Acar 等[17]在 CP 分解的基础上提出了 CP- wopt, 并采用梯度下降算法对模型进行求解。耿氏等人在 Tucker 分解的基础上,提出了将 n 模式降维[19]扩展到不完全张量情况下的 M2SA, 并通过迭代过程得到最终解。Filipović等。[20]提出 WTucker 模型是解决非线性共轭梯度(NCG)方法。最近, Wu 等人[22]基于 Cauchy 分布提出了一种鲁棒的低秩张量分解框架, 并分别开发了基于 Tucker 和 CP 的 CTD 和 CCD 算法。这些方法都是基于张量因式分解, 通常在一定范围内选取精确的张量秩。当选择的张量足够小时, 恢复的张量将具有低秩。但是张量的实秩是不容易估计的。

另一种提取低秩张量的方法是利用传统的矩阵补全方法。例如，核范数最初是一个来自矩阵分析的概念。证明了矩阵秩范数[29]的最紧凸替代。因此，矩阵秩的最小化可以变换为核范数的最小化，核范数的最小化比秩范数的最小化更容易处理。据我们所知，Liu[23]首先将矩阵核范数推广到张量情况下，然后利用张量核范数正则化使张量秩压缩。该方法以每一种方式展开张量，并对展开矩阵的核范数进行惩罚。最后用 ADMM 求解最优问题。在[23]的框架下，在[24][25][26][27]中开发了一些模型。其中，TNCP[27]利用矩阵秩小于因子矩阵的不等式，成功地将 CP 因子分解引入到框架中。

塔克分解方法通常采用交替最小二乘(ALS)算法 HOOI[30]求解，是一个迭代过程。作为 Tucker 的非删节版本，HOSVD[16]遵循 HOOI 但没有迭代，因此比 Tucker 花费的时间要少得多。目前，通过简单地将 HOSVD 的因子矩阵限制为其前导列，截断后的 HOSVD (T-HOSVD) 得到了 Tucker 分解的一个很好的近似。为了获得更高的精度，Vannieuwenhoven[1]提出了一种新的 HOSVD 截断方法，称为顺序截断 HOSVD (ST-HOSVD)，该方法采用贪婪的方法依次得到因子矩阵。由于贪婪的策略，它相当快和惊人的准确性。与 Tucker、CP 和 T-HOSVD 等张量分解方法相比，该方法的另一个优点是可以方便地给出其误差分析。

考虑到大多数张量分解方法都需要指定张量秩作为输入，很难给出最佳张量秩，本文首先提出了一种基于 ST-HOSVD 的自适应张量分解方法。对于张量秩，我们通过控制一个叫做容错的参数来自适应地限制张量的秩，而不是直接选择秩。利用这种自适应张量因子分解方法，我们得到了一个张量的一系列低秩近似，并取其平均值作为其近似。最后，对于不完全张量，我们提出了一种基于自适应张量分解方法的迭代算法来实现它。感谢 ST-HOSVD，我们的方法在很多情况下都非常快。在实验结果中，我们的方法也获得了良好的精度，我们认为这是由于我们使用平均算子来得到张量的近似。注意，average 操作符是减少全局错误的最流行的方法之一。

修改后：

张量分析作为多路系统的有力工具，已成功应用于计算机视觉[1-4]、音频处理[5]、机器学习[6-11]、神经科学[12]、地球物理学[13]、信号处理[14-15]、数据挖掘[16-20]等领域。张量因子分解在多路数据分析中起着重要的作用。对于完全张量，最重要的张量分解方法有 CANDECOMP/ PARAFAC (CP)分解[21-25]、Tucker 分解[26-29]和高阶奇值分解(HOSVD)[30-33]。在数学上，HOSVD 是 Tucker 因式分解的特殊版本。没有降维，Tucker 分解被简化为 HOSVD。目前，塔克分解通常用交替最小二乘(ALS)算法求解，称为 HOOI[29]。通过简单地将 HOSVD 的因子矩阵限制为其前导列，截断的 HOSVD (T-HOSVD)得到了 Tucker 分解的一个很好的近似。最近，为了获得更好的性能，提出了序列截断 HOSVD (ST-HOSVD)[34]来近似 Tucker 分解，其中因子矩阵是用贪婪的方法顺序得到的。由于贪心的策略，ST-HOSVD 比 HOOI 要快得多，而且精确度很高。此外，ST-HOSVD 与 Tucker、CP 和 T-HOSVD 相比，另一个优点是可以方便地给出近似误差分析。

在实际应用中，数据集中可能会出现缺失值，原因有很多：一个或多个传感器的故障和故障、样本之间不规则的测量间隔或不同传感器的采样频率。近年来，基于缺失数据的张量分析在多路系统中得到了越来越多的关注。我们可以看到，关于这个问题的大多数研究都有一个共同的假设，即原始张量是低秩的。为了得到不完全张量的低秩结构，许多经典的张量分解方法被推广到不完全张量的情况下。基于期望最大化(EM)方法[35]，Bro [36]， Walczak [37]， Tomasi [38]， Acar [39]， Karlsson[40]将 CP 因式分解推广到缺失值。[36]、[37]、[38]、[39]、[40]分别使用交替最小二乘法(ALS)、Levenberg-Marquadt 方法[41]、梯度下降算法和并行算法更新因子矩阵。最近，Shao[42]提出了一种基于 CP 因子分解的多源不完整数据聚类稀疏模型。Yokota[43]针对缺失率极高的情况，提出了一种基于平滑约束和低秩近似的有效

平滑 CP 因子分解方法。Zhao[44]提出了一种基于贝叶斯推理和 CP 因子分解的生成模型来处理缺失数据和离群值。基于[36-39]的思想，耿[45]对不完全张量数据提出了一种基于 Tucker 分解的多线性子空间分析(MSA)方法，并将其应用于人脸识别和人脸年龄估计。除了上述基于张量因子分解的方法外，还利用张量核范数来恢复不完全张量的低秩结构。以往的研究表明，核范数是矩阵秩范数[48]的最紧凸替代。在此基础上，Liu[49]将矩阵核范数推广到张量情况，利用张量核范数正则化使张量秩收缩。按照[49]中的框架，Gandy[50]处理了带有噪声的张量补全问题，Liu[51]提出了一个更一般的张量核范数定义，并将其应用于低秩张量补全问题。[51]的主要缺点是该算法涉及很多非常大矩阵的 svd，因此计算代价非常高。为了克服这一缺点，Liu[52]利用矩阵的秩小于因子矩阵的秩的不等式，将 CP 因子分解引入到该框架中。

本文提出了一种处理张量补全的 ST-HOSVD 算法。考虑到秩不便于指定，本文首先提出了一种基于 ST-HOSVD 的完全张量分解自适应方法。对于张量秩，我们通过控制一个容差参数自适应地限制它，而不是直接选择秩。这种自适应张量因式分解法，生成了张量的一系列低秩近似，并取其平均值作为其低秩近似。最后，对于不完全张量，我们提出了一种基于自适应 ST-HOSVD 和平均算子的迭代算法。在图像和视频上的实验结果表明，与其它不完全张量算法相比，该方法具有较高的计算速度和较高的精度。

[返回](#)

附件二：实验和分析的示例

In this section, we evaluate the proposed method on 14 colorful image datasets and 3 video datasets. All experiments are run on MATLAB R2015a on an Intel i7-4790 3.60 GHz machine with 12GB memory. Given a tensor with missing data, the relative square error (RSE) of the recovered tensor is given by RSE.

We compare our method with Trace Norm Regularized CP (TNCP) [52], CP Weighted OPTimization (CP-WOPT) [39], WTucker [46], Multilinear Subspace Analysis with Missing Data (M2SA) [45] and Cauchy Tucker Decomposition (CTD) [47].

Experiments on Colorful Images Data

Experiments on Images with Noise

Experiments on Images with Different Sizes

Experiments on 4D Video Data

Comparing with T-HOSVD based Methods

Optimal Parameters

翻译版本

在本节中，我们对 14 个彩色图像数据集和 3 个视频数据集进行了评估。所有实验在 Intel i7-4790 3.60 GHz、12GB 内存的机器上运行于 MATLAB R2015a 上。给出了一个数据缺失张量的相对平方误差（RSE）。

（1）我们的方法与跟踪范数正则化 CP（TNCP）[52]、CP 加权优化 CP（CP-wopt）[39]、

WTucker[46]、缺失数据多线性子空间分析 M2SA[45]和柯西塔克分解 (CTD) [47]进行了比较。

- (2) 彩色图像数据实验
- (3) 噪声图像实验
- (4) 对不同大小的图像进行实验
- (5) 4D 视频数据实验
- (6) 与基于 T-HOSVD 的方法进行比较
- (7) 最优参数

[返回](#)

附件三：结论和未来的工作的示例

原始版本：

In this paper, we propose a method based on ST-HOSVD for the recovering of incomplete tensors. Unlike the other tensor factorization method, we try a new strategy to capture the low-rank structure of a tensor by the error tolerance parameter instead of specifying the exact tensor rank. With this strategy, an iterative algorithm is proposed for incomplete tensors. The experiments show that our method can recover the incomplete tensors with a good accuracy. The advantage of our method is its speed, often times faster than the other methods.

修改后的版本：

In this paper, an iterative algorithm is proposed based on adaptive ST-HOSVD and the average operator of low-rank approximation for recovering incomplete tensors. Unlike the other tensor factorization based methods, we capture the low-rank structure of a tensor by the error tolerance parameter instead of specifying the exact tensor rank. With this strategy, the average operator is used to obtain a better approximation of a tensor. Theoretical analysis shows that the convergence of the proposed algorithm can be guaranteed. The experimental results on images and videos show that in most cases, the proposed method has the best accuracy and the shortest running time compared with the state-of-the-art methods for recovering the incomplete tensors.

Along the line of the present work, some problems deserve further research. For example, some acceleration technologies can be studied for speeding up the proposed algorithm. Based on the fact that HOSVD has denoising function for images and videos, we will be interested to apply the proposed algorithm to tensor data with noise. In the future work, it is very interesting to apply the proposed algorithm to the relational data in social network and multi-view learning.

翻译版本

原始版本：

本文提出了一种基于 ST-HOSVD 的不完全张量恢复方法。与其他张量因子分解方法不同，我们尝试了一种新的策略，通过误差容限参数来获取张量的低秩结构，而不是指定张量的精确秩。利用该策略，提出了一种求解不完全张量的迭代算法。实验表明，该方法能较好地恢复不完全张量。我们的方法的优点是它的速度，通常比其他方法快好几倍。

修正版:

本文提出了一种基于自适应 ST-HOSVD 和低秩近似平均算子的不完全张量恢复迭代算法。与其它基于张量因子分解的方法不同，我们通过误差容限参数来获取张量的低秩结构，而不是指定张量的精确秩。用这种策略，平均算子被用来获得一个张量的更好的近似。理论分析表明，该算法能保证算法的收敛性。在图像和视频上的实验结果表明，在大多数情况下，与目前的不完全张量恢复方法相比，该方法具有最好的精度和最短的运行时间。(修正版)

在目前工作的基础上，有一些问题值得进一步研究。例如，可以研究一些加速技术来加速提出的算法。基于 HOSVD 对图像和视频具有去噪功能，我们将有兴趣将所提出的算法应用于有噪声的张量数据。在未来的工作中，将该算法应用于社交网络中的关系数据和多视角学习是非常有趣的。

[返回](#)

附件四：一个摘要修改的例子

原摘要:

Abstract—The problem of recovering missing data of an incomplete tensor has drawn more and more attention in many fields such as machine learning, signal processing and so on. Researches on this problem usually share a common assumption that the original tensor is of low-rank. There has been two ways to capture the low-rank structure of the incomplete tensor. One is based on tensor factorization, while the other extends the matrix completion methods. Recently, a novel tensor factorization method called ST-HOSVD is presented with high speed and good accuracy. In this paper, we present an algorithm on tensor completion. For traditional factorization algorithms, tensor ranks should be specified ahead, which we think is not reasonable. Thus we propose our strategy for tensor factorization based on ST-HOSVD without need of tensor ranks. The only parameter required here is the error tolerance, making it easy to tune the parameters. Our method absorbs the advantage of ST-HOSVD and it performs very fast. Following the progress of ST-HOSVD, we present a novel way using the average operator, which is quite different to existing methods, to improve the accuracy of the fulfilled tensor. Finally, in the experiments our method outperforms the others in both running time and the accuracy, which proves our judgement.

修改后:

Abstract—The problem of recovering missing data of an incomplete tensor has drawn more and more attentions in the fields of pattern recognition, machine learning, data mining, computer vision and signal processing. Researches on this problem usually share a common assumption that the original tensor is of low-rank. One of the important ways to capture the low-rank structure of the incomplete tensor is based on tensor factorization. For the traditional tensor factorization algorithms, the tensor ranks should be specified ahead, which is not reasonable in real applications. To overcome this drawback, an adaptive algorithm is first presented based on sequentially truncated higher order singular value decomposition (ST-HOSVD) for fast low-rank approximation of complete tensor, in which the tensor ranks can be obtained adaptively. Then for tensor with missing data, we use adaptive ST-HOSVD and the average

operator of low-rank approximation to improve the accuracy of the fulfilled tensor. Convergence analysis of the proposed algorithm is also given in this study. The experimental results on 14 image datasets and 3 video datasets show that the proposed method outperforms the state-of-the-art methods in terms of running time and the accuracy.

翻译版本

原摘要:

摘要---恢复不完整张量的缺失数据的问题在机器学习, 信号处理等许多领域引起了越来越多的关注。关于这个问题的研究通常有一个共同的假设, 即原始张量是低秩的。捕获不完整张量的低秩结构有两种方法。一种基于张量分解, 而另一种则扩展了矩阵完成方法。最近, 提出了一种新的张量分解方法, 称为 ST-HOSVD, 具有很高的速度和较高的精度。在本文中, 我们提出了一种关于张量完成的算法。对于传统的因式分解算法, 应提前指定张量等级, 我们认为这是不合理的。因此, 我们提出了基于 ST-HOSVD 的张量分解策略, 而无需张量等级。此处所需的唯一参数是容错能力, 可以轻松调整参数。我们的方法吸收了 ST-HOSVD 的优势, 并且执行速度非常快。随着 ST-HOSVD 的发展, 我们提出了一种使用平均算子的新颖方法, 该方法与现有方法完全不同, 可以提高完成张量的精度。最后, 在实验中, 我们的方法在运行时间和准确性上都优于其他方法, 这证明了我们的判断。

修改后:

摘要---恢复不完整张量的缺失数据的问题在模式识别, 机器学习, 数据挖掘, 计算机视觉和信号处理领域引起了越来越多的关注。关于这个问题的研究通常有一个共同的假设, 即原始张量是低秩的。捕获不完全张量的低秩结构的重要方法之一是基于张量分解。对于传统的张量分解算法, 应预先指定张量等级, 这在实际应用中是不合理的。为了克服这个缺点, 首先提出了一种基于顺序截断的高阶奇异值分解 (ST-HOSVD) 的自适应算法, 用于快速完成完整张量的低秩逼近, 其中自适应地获得张量等级。然后对于缺少数据的张量, 我们使用自适应 ST-HOSVD 和低秩逼近的平均算子来提高完成张量的精度。本文还对算法进行了收敛性分析。在 14 个图像数据集和 3 个视频数据集上的实验结果表明, 该方法在运行时间和准确性方面都优于最新方法。

[返回](#)

附件五：一个同行评议的例子

同行评议过程的例子来自以下公认的期刊论文:

Chen Sentao, Han Le, Liu Xiaolan, He Zongyao, Yang Xiaowei, Subspace Distribution Adaptation Frameworks for Domain Adaptation, IEEE Transactions on Neural Networks and Learning Systems, 2020. DOI: 10.1109/ TNNLS.2020.2964790.

副主编的第一次决定

对作者的评论:

审稿人提出了一些关于实验和论文贡献的问题。特别是, 本文最大的问题是该方法适用于低维特征空间。虽然这可能对一些低维任务有效, 但作者需要展示如何将其转换为高维任

务。此外，还必须包括额外的实验和分析。

副主编的第二次决定

对作者的评论：

已收到对该文件进行这一修订的报告。

(1) 本文最大的问题是该方法适用于低维特征空间。作者增加了新的数字分类和人脸识别实验，但仍在低维。作者应该讨论/展示如何将其应用于大型自然图像。

(2) 需要与深度学习方法进行比较。鼓励作者考虑审稿人的反馈意见，编写第二次修订版。在这样做的时候，还应该准备好对审稿者的审稿意见的逐点回答。

请仔细校对并改进语言表达。作者被鼓励找一个英语为母语的人或使用一个专业的编辑服务（更多信息见 TNNLS 网站，“作者信息”下的“专业编辑服务”一节）来帮助改善语言表达

作者应该清楚地说明对 NNLS 核心的重要的原始贡献。TNNLS 的范围是指“神经网络和相关学习系统的理论、设计和应用”。因此，每一次提交的报告都应神经网络和相关学习系统的核心有明确而牢固的联系。请明确说明这一点。作者也被鼓励去看看最近在 TNNLS 上发表的关于这个主题的论文，以显示与 NNLS 社区的紧密联系。

审稿者的意见

审核人 1 对作者的评论：

作者们已经充分地阐述了我对原稿微调的评论。因此，在我看来，这篇论文现在可以发表了。

审核人 2 对作者的评论：

上一次修订解决了我所有的顾虑，我建议接受这篇论文。作者被鼓励（不一定）添加更详细的解释（p16, I53-59），以便在相关工作中与 DL 方法（训练示例、训练时间、模型大小、运行时间等）进行比较。

审核人 3 对作者的评论：

我读了作者的反驳，这让我对这篇文章有了更深的理解。本文符合验收标准。

副主编的第三次决定

对作者的评论：你是如何在上一轮中完全解决副主编的所有意见，阐述的还不清楚。请参考副主编的上一轮评审意见，并严格处理上一轮提出的所有意见和建议，包括语言问题。

主编的决定

尊敬的陈先生：

您提交给 IEEE 神经网络和学习系统事务的手稿 TNNLS-2018-P-10341.R2，题为“Subspace Distribution Adaptation Frameworks for Domain Adaptation”，已经过审查。

审核人的意见包含在本电子邮件的底部。我很高兴地通知你，你的论文已收到积极的建议，可在有条件的情况下发表，但须经评论中指出的修改。审稿人和副主编的意见和建议应当严格执行。

我邀请你回复审稿人的意见并修改你的手稿。修改后的稿件将由总编辑审阅，可能还将由副主编和一些审稿人审阅。请注意，如果您没有完全处理审稿人和副主编的所有意见，您的论文仍有可能被拒绝。

[返回](#)

