

目录

前言	1
一、文献翻译软件	1
二、读英文文献的方法	1
2.1 第一步阅读 Abstract	2
2.2 第二步阅读 Introduction	2
2.3 第三步阅读 Conclusion	3
2.4.第四步阅读 Table	3
三、笔记格式	5
3.1 主旨笔记	5
3.2 研究方法笔记	5
3.3 其他笔记	6

前言

作为理工科的学生，我们经常需要阅读英文文献甚至动笔写作来交流学习前沿的科学技术。但是在对于一个领域不够了解，而且英语水平稍有欠缺的情况下，从头到尾按顺序阅读是非常费劲的。同时，对于阅读英文文献时如何做出有条理的笔记方便日后复习也是非常值得探讨的问题。因此，我在网上查阅整理了有关阅读英文文献的有效方法，并以一篇最近阅读过的降维领域的英文文献“2-D_Embedding_of_Large_and_High-dimensional_Data_”作为例子说明，整理汇总如下文。

一、文献翻译软件

文献中往往有大量的专业术语和图片表格，将英文文献整篇地复制进百度翻译等翻译软件是不现实的也不方便的。因此在这里我使用的是“知云文献翻译”软件。这个软件是免费使用的，可以翻译 PDF 和 word 版的文献，而且页面简洁，导入文件之后页面左边是文献原文，拖动需要翻译的部分之后右边就会出现相应的中文翻译。它有多种翻译引擎供我们选择。同时它还提供对于文献做注释的功能，非常方便我们阅读过程中做出标识。

经过一段时间的使用，我认为该软件操作简单，翻译准确度较高，很适合通篇泛读时快速理解文章意思。

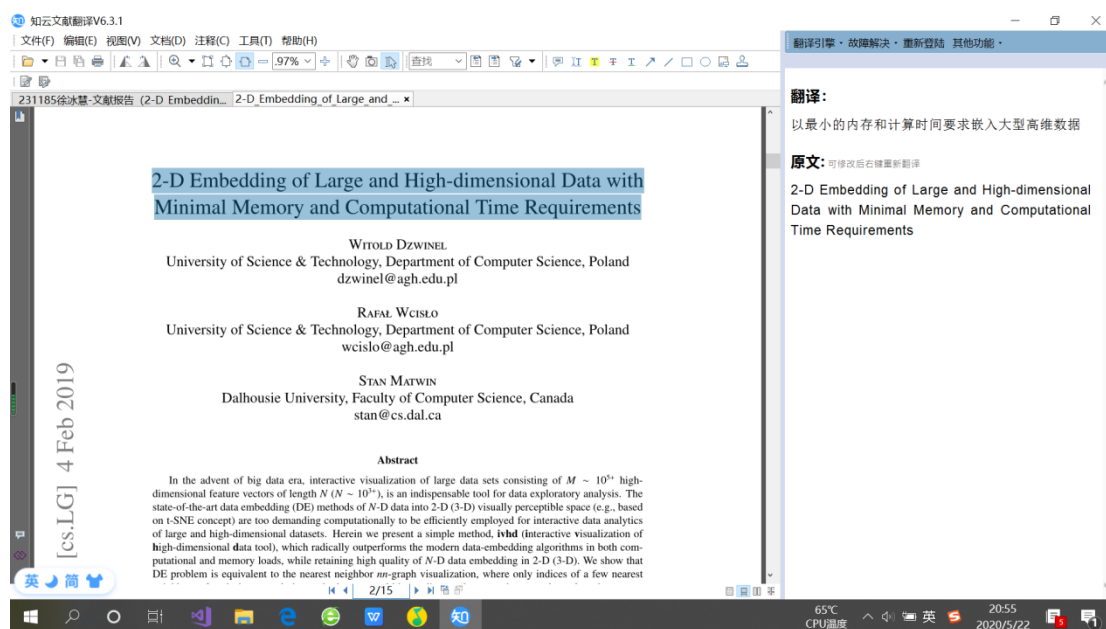


图 1-1 知云文献翻译的界面

二、阅读文献的顺序

2.1 阅读前的准备

了解自己的需求：

- 1) 了解领域动态。那你看看摘要就好了，读懂标题，知道本文的核心结论就好。具体要求见知乎文章。
- 2) 准备入一行，扩充知识面。重点读引言，看图表结果，大致了解本文的逻辑推理路线。
- 3) 准备借鉴此文开展自己的研究工作。这就是最深层次的阅读了，必须要理解本文的设计理念，此时讨论就尤为重要。

误区：

- 1) 不要高亮不认识的词，没意义，只能高亮文章的逻辑核心词汇和语句。
- 2) 不要一开始就通读，这样你很快就读不下去；正确的做法是由总到细，由浅及深，分步阅读。
- 3) 不要读完就不管了，在没有自我总结之前，知识是不属于你的
- 4) 不要以为文献读一遍就够了，理解认知是一个螺旋循环上升的过程。

2.1 第一步阅读 Abstract

摘要本身就是对全篇内容的概括，是重中之重。一篇论文的核心往往是一两句话，而论文的其他部分都是在用不同的方式支持它的核心，因此读好摘要是非常重要的。

比如如图 2-1 我们可以对摘要进行这样的注释，从而对它进行分解。

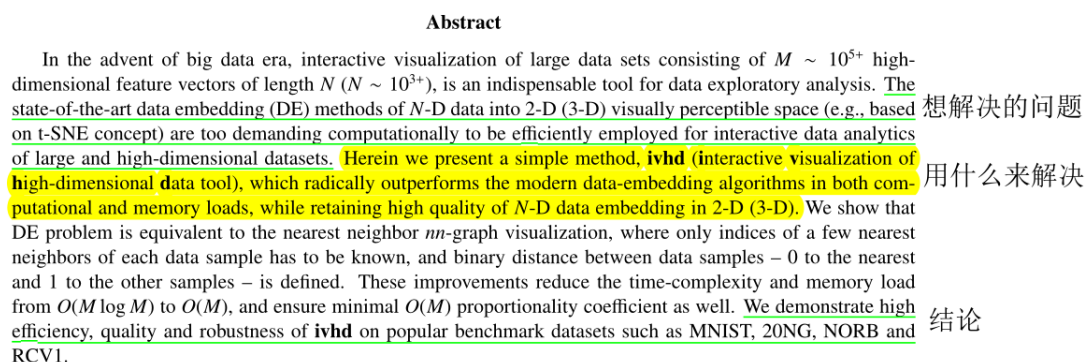


图 2-1 对摘要的注释

Abstract 中我们应该需要做的笔记：

- ①作者想解决什么问题？ **question**
- ②作者通过什么理论/模型来解决这个问题？ **method**
- ③作者给出的结论是什么？ **Answer**

通过摘要，可以得到的信息有：

- 1.这篇文章与我的研究内容是否相关？
- 2.采用的方法是否新颖，我是否感兴趣？
- 3.他的结果，是我想要的么？

以上三个问题，可以变化你阅读论文心中所带的问题，即：

- 1.他是怎么把这种方法运用到这类问题中的？（学习）
- 2.它的结果完美么？是否是充分必要的？保守性强不强？有没有可以改进的地方？（总结+反思）

所以，阅读论文的 **Abstract**，实际是你为阅读论文所做的最重要的准备工作，带着问题去读论文才最有效率，万万不可跳过 **Abstract**。

PS：每次读完文献（不管是细读还是粗读），合上文献后，想想看，文章最重要的 **take home message** 是什么，如果不知道，就从 **abstract, conclusion** 里找，并且从 **discuss** 里最好确认一下。这样一来，一篇文章就过关了。**take home message** 其实都不会很多，基本上是一些 **concepts**，如果你发现你需要记得很多，那往往是没有读到重点。

2.2 第二步阅读 **Introduction**

这一部分其实是对于**文章的研究背景和研究意义**的介绍，有些论文也会将文献综述的简化版加在里面。在 **Introduction** 的最后，作者会介绍文章的架构，让你清楚文章的每一部分都在做什么。

Introduction 中我们需要做的笔记：

- ①此问题的由来（历史），作者为什么研究这个课题？
- ②此问题在全世界内的研究进展（现状），目前这个课题的研究进行到了哪一阶段？

③引申出作者的研究动机，作者使用的理论是基于哪些假设？

Tips: introduction 可以读完，泛读时着重读前两段。此时你必须能回答出几个问题才算真正抓住了本文的核心：文章的引言、结果和结论是如何将摘要详细展开的。

2.3 第三步阅读 Conclusion

Conclusion 通常会把整个文章的主要内容复述一遍，帮助读者回顾+理清思路，然后在此基础上深入自己的研究。

在 Abstract 或者 Introduction 中，作者就已经给出了课题的结论。而在 Conclusion 中会评价自己的结论，并且在这个课题上的基础上**进一步探讨**，比如延伸到这个课题未来还有哪些研究方向，可以用来解决哪些问题，等等。

Conclusion 中我们需要做的笔记：

①这篇文章存在哪些缺陷？

②作者关于这个课题的构思有哪几点？

读完以上三步，并做完笔记以后，我们已经对这篇文章的**核心**有了比较清晰的认识。接下来需要了解的就是**作者如何像做数学证明题一样，一步一步得出了他的结论**。

但是如果直接从研究方法 method 开始读，复杂的公式往往让人很难有耐心地进行下去，因此**要搞清楚作者的研究方法，应该先看的不是理论的推导过程，而是文章中的表格**。

2.4.第四步阅读 Table

在阅读完前面的三个部分，我们对于文章的核心已经有了一个把握。接下来我们要了解作者得到结论的方法。但如果我们按顺序从 background 或者 key concept 阅读，其中复杂的术语和不知所云的公式往往会让我们很快放弃。例如本篇文献对于刚接触降维以及可视化的我来说，在不了解类似图 2-2 中的复杂公式与指标与文章有什么联系的前提下，是很难有耐心查阅所有专业术语逐行理解的。

To obtain 2-D embedding \mathbf{X} of $nnG(\mathbf{V}, \mathbf{E})$ we minimize the following stress function:

$$E(\|\mathbf{D} - \mathbf{d}\|) = \sum_i \sum_{j \in O_m(i) \cup O_m(i)} b(\delta_{ij} - d_{ij})^2. \quad (9)$$

Consequently, the interparticle force \mathbf{f}_i from Eq. (3) in $E(\cdot)$ minimization procedure simplifies to:

$$\mathbf{f}_i^n = -\sum_{j \in O_m(i)} \mathbf{x}_{ij}^n - c \sum_{k \in O_m(i)} (1 - d_{ik}^n) \cdot \frac{\mathbf{x}_{ik}^n}{d_{ik}^n}, \quad \mathbf{x}_{ik}^n = \mathbf{x}_i^n - \mathbf{x}_k^n. \quad (10)$$

图 2-2 文献中的复杂公式

因此我们此时应该先阅读表格。因为表格给出的信息会简单直观很多，并且它们通常都是作者研究过程中某一阶段成果的总结。

例如本篇文献中的 Table 1（图 2-3），从表格标题就可以看出这一步作者的操作是：为初始化和 nn 近邻搜索选择的计时。如果对于这一步的目的、操作或者结论有所疑问，也可以马上在表格上下的解释中获得答案（图 2-4）。这个方法对于我们把握作者的研究思路和步骤是非常好的。

Table 1: Selected timings for initialization and nn nearest neighbors search.

Data set	Brute-force	Approx.
MNIST ($M = 7 \cdot 10^4$, $N = 30$, $nn = 100$)	1m 27s	1m 35s
MNIST ($M = 7 \cdot 10^4$, $N = 30$, $nn = 2$)	1m 17s	39s
MNIST ($M = 7 \cdot 10^4$, $N = 784$, $nn = 2$)	20m 16s	2m 49s
MNIST ($M = 7 \cdot 10^4$, $N = 784$, $nn = 100$)	20m 29s	32m 06s

图 2-3 Table 1

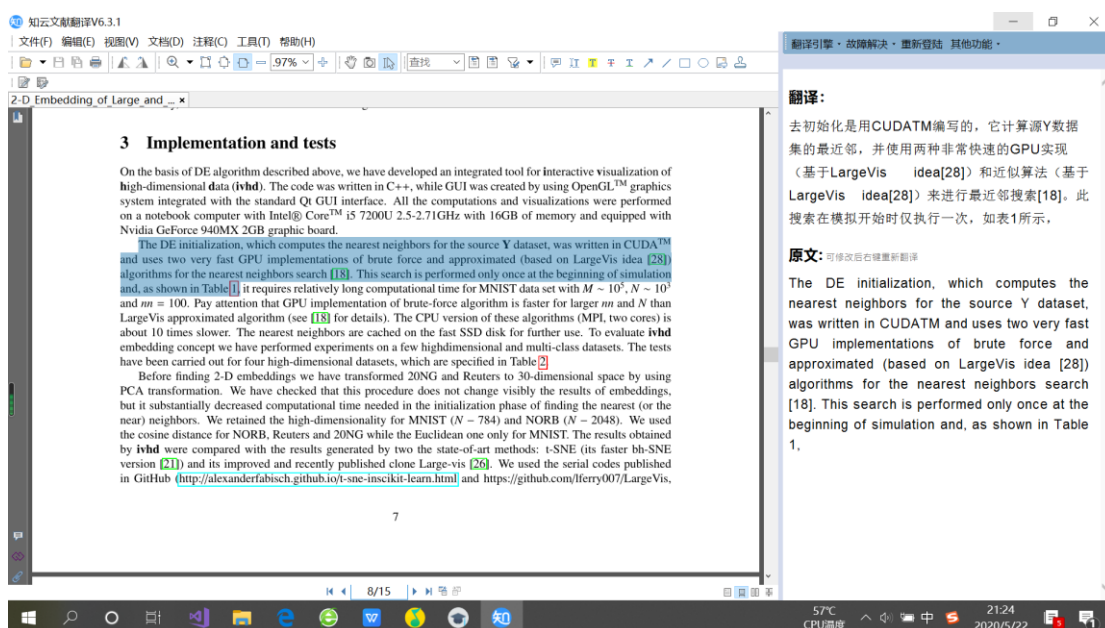


图 2-4 对于 Table 1 的解释

表中清晰地展现了已经获取的变量和得出的结果。表格下面的 **Notes**，则清晰地给出了这个表格研究的对象+模型用到的数据。

而正文中对表格进行阐述的那一段则向我们解释了这个表格中的数据是什么，通过什么模型、理论，得出了什么样的结论，而这个结论又有哪些不足，文章还需要进一步探讨什么。

关于数据的来源，几乎所有做 **empirical research**（实证研究）的文章都会给出。这一段会介绍数据都是从哪些数据库扒下来的，采用的哪个时期的数据，数据包含哪些指标，文章后面提到的一些指标和变量具体是如何计算的。

因为都是对数据的解释，所以只需要在整个文章段落中定位你所要查询的变量名词，就能快速了解这个变量从收集到计算的过程。

阅读 Table 需要做的笔记：

- ①研究的数据从哪里来？
- ② 研究中用到的重要指标有哪些？
- ③ 模型分哪几步？每一步分别得出了什么结论？

方法和结果部分读起来会有这样的感觉：读到一半，居然忘了自己在看什么，忘了自己为什么要读这篇文章，忘了他到底在证明些什么东西，还有就是每个环节都懂了，最后连在一起还是不懂...那么说明你的大脑已经跟不上伟人的脚步了，你需要一杆笔和一张纸。

读一点，就在纸上做一点笔记，有问题标记出来，读完一遍之后回头看这些问题，或去查资料，或去请教老师，然后再读一遍论文，一篇好的论文，值得你去来回读个至少五遍的。

三、笔记提纲

根据上面对于文献的剖析，可以按以下提纲进行文献的记录。这样的笔记让我们对整篇文章的脉络都有非常清晰的认识，同时这样的积累对于我们把握这个领域的知识和提高阅读文献的速度都很有好处。好的笔记不仅让我们积累知识，同时也方便我们日后复习到论文中的重要信息。

3.1 主旨笔记：

(1)作者想解决什么问题？ question

(2)作者通过什么理论/模型来解决这个问题？ method

- (3) 作者给出的答案是什么？ answer
- (4) 作者为什么研究这个课题？
- (5) 目前这个课题的研究进行到了哪一阶段？
- (6) 作者使用理论基于哪些假设？
- (7) 这篇文章存在哪些缺陷？
- (8) 作者关于这个课题的构思有哪几点？

3.2 研究方法笔记：

- (1) 研究的数据从哪里来？
- (2) 研究中用到的重要指标有哪些？
- (3) 模型分哪几步？每一步分别得出了什么结论？

3.3 其他笔记

- (1) 记录该领域常用专业术语和论文表达。
- (2) 对于不懂的理论名称，模型名称，专业名词在网上查询并做好记录。

文献阅读几点启示：

1. 每月关注《IEEE Evolutionary Computation》和《Evolutionary Computation》，定期查看《科学》和《自然》。如果时间有限，读读文章标题以及与自己研究领域相关的科研论文的 abstract，即可！这样做可以保证一个学生基本上能够跟踪本领域最重要的发现和进展，同时开阔视野，大概知道其它领域的动态。顶刊上的综述文献往往都是本领域的牛人们写的。对此要精读，要分析其文章的构架，特别要关于作者对各个方向的优缺点的评价以及对缺点的改进和展望。通过精读一篇好的英文综述，所获得的不只是对本领域现在发展状况的了解，同时也可以学会很多地道的英文表达。
2. 在读具体的科研论文时，最重要的是了解文章的主线逻辑。文章中的所有 Figures 都是按照这个主线逻辑展开描述的。所以，一般先读“introduction”部分，然后很快地看一遍 Figures。大概知道这条主线之后，才一字一句地去读“results”和“discussion”。
3. 当遇到一些实验或结果分析很晦涩难懂时，不必花太多时间深究，而力求一气把文章读完。也许你的问题在后面的内容中自然就有解答。这与听学术讲座非常相似！你如果想每个细节都听懂，留心每一个技术细节，那你听学术讲座不仅会很累，而且也许会为了深究一个小技术环节而影响了对整个讲座逻辑推理及核心结论的理解。
4. 对个别重要的文章和自己领域内的科研论文，应该精读。对与自己课题相关的每一篇论

文则必须字斟句酌地读。这些论文，不仅要完全读懂，理解每一个实验的细节、分析、结论，还必须联想到这些实验和结论对自己的课题的影响和启发，提出自己的观点。

5. 科学论文的阅读水平是循序渐进的。每个人开始都会很吃力，所以你有这种感觉不要气馁。坚持很重要，你一定会渐入佳境。当你有问题时或有绝妙分析时，应该与师兄师姐或找导师讨论。

6. 科研训练的一个重要组成部分就是科研论文的阅读。每一个博士生必须经过严格的科研论文阅读的训练。除了你自己的习惯性阅读外，你应该在研究生阶段选修以阅读分析专业文献为主的一至两门课，在实验室内也要有定期的科研论文讨论（Journal Club）。如果你的实验室还没有这种讨论，你们学生可以自发地组织起来。

7. 前面几条都是讨论如何提高科研论文的阅读能力，但是一旦入了门，就要学会 **critical reading**。不要迷信已发表的论文，哪怕是发表在非常好的期刊上。要时刻提醒自己：该论文逻辑是否严谨，数据是否可靠，实验证据是否支持结论，你是否能想出更好的实验，你是否可以在此论文的基础上提出新的重要问题？等等。

天外有天，读科研论文是一件很简单、但也很深奥的事情。一般的学生常常满足于读懂、读透一篇好的论文，优秀的学生则会举一反三、通过查找 **references** 纵深了解整个领域的历史、现状，并展望该领域未来的可能进展。

最后就是针对自己的课题查找阅读相关英文文献了。以下四类英文文献是我们所需要的：

1. 本领域核心期刊的文献。《IEEE Evolutionary Computation》和《Evolutionary Computation》
2. 本领域牛人或者主要课题组的文献。每个领域都有几个所谓的领军人物，他们所从事的方向往往代表目前的发展主流。因此阅读这些组里的文献就可以把握目前的研究重点。
3. 高引用次数的文章。一般来说高引用次数（如果不是靠自引堆上去的话）文章都是比较经典的文章，要么思路比较好，要么材料性能比较好，同时其文笔应该也不赖的话。多读这样的文章，体会作者对文章结构的把握和图表分析的处理，相信可以从中领悟很多东西的。
4. 最后就是当你有了一定背景知识，开始做实验并准备写论文的时候需要看的文献了。首先要明确一点，你所做的实验想解决什么问题？明确这一点后，就可以有的放矢查找你需要的文献了。而且往往当你找到一篇与你研究方向相近的文章后，通过 **ISI** 的反查，你可以找到引用它的文献和它引用的文献，从而建立一个文献树，更多的获取信息量。（要明白自己试验的创新，这一点是科研的关键！！）

231185 徐冰慧