

基于 Python 的矩阵转置方法的分析与研究

王莉丽¹

WANG Lili

摘要

为深入研究 Python 中矩阵转置的方法, 文章系统分析了 NumPy 库中 T 转置、Transpose 方法和 Swapaxes 方法的特性。针对 Transpose 方法, 提出基于多维坐标轴的图形表示法以辅助理解其原理; 同时, 梳理当前矩阵转置算法在不同领域的应用现状, 并基于二维矩阵经典算法, 设计了一种适用于多维矩阵的新型转置算法。结果表明, 采用坐标轴图形表示法能够直观展现数据变换过程; 新设计的多维矩阵转置简易算法可快速完成转置操作, 既有助于深入理解转置逻辑, 又为高效实现矩阵转置提供了新路径。未来需进一步加强对高维矩阵存储方法及其适用硬件平台的研究。

关键词

Python; NumPy; 矩阵转置; 多维矩阵; 转置算法

doi: 10.3969/j.issn.1672-9528.2025.05.042

0 引言

Python 第三方库 NumPy 主要用来存储和处理大型矩阵, 支持多维数组和矩阵运算。在线性代数、多维图像和信号处理中, 矩阵转置是基本操作; 转置研究对图形变换也具有几何意义^[1]。文章对 NumPy 中进行矩阵转置的三种方法进行分析, 并就 Transpose 方法提出一种多维坐标轴的图形表示方法, 以期帮助研究者更好地理解该方法。

1 矩阵转置方法分析

矩阵转置是指将矩阵的行列互换, 得到一个新的矩阵。NumPy 完成矩阵转置的方法有 3 种。

1.1 T 转置

T 转置对矩阵的行、列进行互换, 常用于二维矩阵。

例如, 一个 $n \times m$ 的矩阵 `arr`, 通过 T 转置的操作, 可得到 $m \times n$ 的矩阵。对矩阵中任一数据, 行列索引互换可得到转置后的位置索引, 如 `arr[i][j]=arr.T[j][i]`。T 转置对一维矩阵无意义, 不会改变矩阵内容。

多维数组的 T 转置相当于把轴编号逆序的结果。如一个四维矩阵 `data`, 形状是 (2, 3, 4, 5)。多维矩阵的轴编号从 0 开始, 四维矩阵 `data` 四个轴编号分别为 0、1、2、3; `data.T` 等价于 `data.transpose(3, 2, 1, 0)`, 转置后的矩阵形状为 (5, 4, 3, 2)。

多维矩阵中的任一数据 `data[a][b][c][d]`, 通过 T 转置后的下标为轴逆序的结果, 即 `data[a][b][c][d]=data.T[d][c][b][a]`。

多维矩阵的 T 转置结果单一, 因此对多维矩阵的转置操作一般有两种方法。

1.2 Transpose 方法

Transpose 方法可同时交换两个轴或多个轴索引, 常用于多维矩阵。Transpose 方法对于二维矩阵的操作等价于 T 转置, 只交换行列索引; 对于多维矩阵的轴转置操作更灵活, 可以在参数中指定任意顺序的轴编号。

从 Transpose 参数中轴排列的顺序可以直观地得到转置后轴索引的规律。如四维矩阵 `data`, `data.transpose(0, 1, 2, 3)` 参数中轴编号顺序未变, 因而未对矩阵做任何改变; `data.transpose(3, 0, 1, 2)` 的轴变换方式如图 1 所示。

`data.transpose(3, 0, 1, 2)`
轴正常编号顺序: 0 1 2 3
转置后: `data` 原 3 轴现转换为 0 轴
`data` 原 0 轴现转换为 1 轴
`data` 原 1 轴现转换为 2 轴
`data` 原 2 轴现转换为 3 轴

图 1 Transpose 轴变换图

对于 `data` 中的任一数据, 转置后的索引下标变换和 Transpose 参数中的轴变换方式相同。比如 `data[a][b][c][d]=data.transpose(3,0,1,2)[d][a][b][c]`。

1.3 Swapaxes 方法

Swapaxes 参数只能接收两个坐标轴编号, 即一次只能交换两个轴上的索引, 可以看作简化版的 Transpose。

如四维矩阵 `data`, `data.swapaxes(2,3)` 等价于 `data.transpose(0,1,3,2)`, 只交换 2 轴和 3 轴的索引; 转置后的

1. 河南工业贸易职业学院 河南郑州 450003

数据索引下标也只转换原矩阵中 2 轴和 3 轴的元素索引，
`data.swapaxes(2,3)[a][b][c][d]=data.transpose(0,1,3,2)[a][b][c][d]=data[a][b][d][c]`。

2 Transpose 方法坐标轴的图形表示

综上 3 种转置方法的分析，可以看出 Transpose 方法适应性广，接下来主要针对该方法研究图形表示方法。

多维矩阵转置中矩阵元素在多个坐标轴中转换比较复杂，为更好地理解数据元素转置后的位置变换，可将多维坐标轴画图表示。

以三维矩阵为例，创建一个三维矩阵 `arr3=np.arange(24).reshape(2, 3, 4)`，`arr3` 的形状为 (2, 3, 4)，即三维矩阵中含有 2 个二维数组，每个二维数组含有 3 个一维数组，每个一维数组含 4 个元素，反映在坐标轴上，即 0 轴有 2 个元素、1 轴有 3 个元素、2 轴有 4 个元素。

基于 Python 语言，创建和显示三维矩阵 `arr3` 的数据信息，代码如图 2 所示。

```
In [1]: import numpy as np
arr3=np.arange(24).reshape(2, 3, 4)
print(arr3)

[[[ 0  1  2  3]
  [ 4  5  6  7]
  [ 8  9 10 11]]

 [[12 13 14 15]
  [16 17 18 19]
  [20 21 22 23]]]
```

图 2 创建和显示 `arr3` 数据的 Python 运行结果图

将 `arr3` 数据标注在坐标轴图上，如图 3 所示。

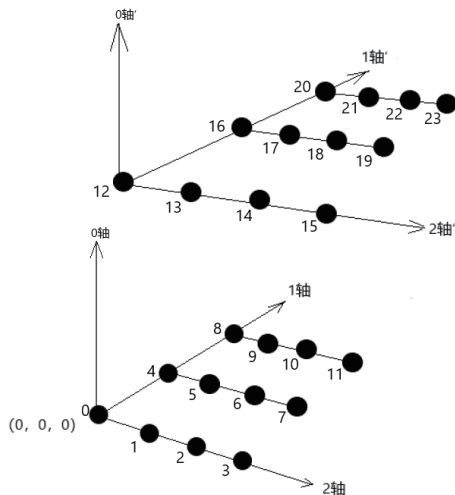


图 3 `arr3` 数据的坐标轴图形表示

需要从 0 轴开始标注，0 轴对应二维数组个数，即第 3 维坐标个数，0 轴上每一个元素对应一个三维平面，可使用 0 轴'标注第 2 个平面，以此类推；1 轴是第 2 维坐标，1 轴

上每个元素对应 1 个一维数组；2 轴数据是 1 维数组元素，用实心点表示数据元素，将 `arr3` 数据标注在坐标轴图上。注意 0 轴保持始终向上的角度，更方便转置后数据的查找。

分析代码中的数组元素和坐标轴图中的表示，可直观地查看数据的分布情况。接下来可先利用坐标轴图形方式，分析 `arr3.transpose(2,0,1)` 转置后数据元素的位置，再使用代码运行结果印证。

`arr3.transpose(2,0,1)` 操作，即转置后的 0 轴、1 轴、2 轴分别是原 2 轴、0 轴、1 轴。转置操作过程中，各数据轴上数据元素不变，3 个坐标轴的交点 (0, 0, 0) 的位置转置过程不会改变，转置后如图 4 所示。

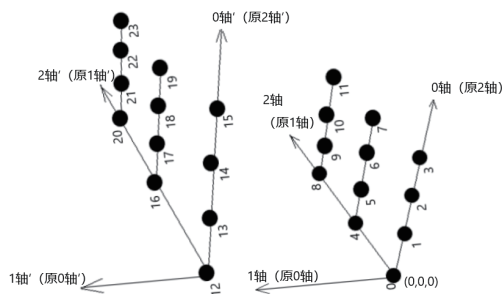


图 4 `arr3` 转置后坐标轴的图形表示

按照每个数据轴代表数组元素特点，即 0 轴（原 2 轴）代表二维数组个数，有 4 个二维数组，数组的框架可表示为：

```
[[[]]
 []]
 []]
 []]
 []];
```

1 轴（原 0 轴）代表一维数组个数，每个二维数组中有 2 个一维数组，数组框架可表示为：

```
[[[] []]
 [] []]
 [] []]
 [] []];
```

2 轴（原 1 轴）代表一维数组中的元素，数据读取顺序先读 0 轴，再读 0' 轴，以此类推。则根据转置后的坐标图将读取到的数据填入数组框架为：

```
[[[0, 4, 8] [12, 16, 20]]
 [[1, 5, 9] [13, 17, 21]]
 [[2, 6, 10] [14, 18, 22]]
 [[3, 7, 11] [15, 19, 23]]];
```

对 `arr3` 进行转置的 Python 代码如图 5 所示，可看到与坐标轴图中分析填入数据一致。利用坐标轴的图形表示方法，可直观地找到三维矩阵的转置数据的变换位置，帮助研究者更快地理解转置过程和数据变换。

```
In [2]: print(arr3.transpose(2,0,1))

[[[ 0  4  8]
  [12 16 20]]

  [[ 1  5  9]
  [13 17 21]]

  [[ 2  6 10]
  [14 18 22]]

  [[ 3  7 11]
  [15 19 23]]]
```

图5 arr3 矩阵转置的 Python 运行结果图

3 一种多维矩阵转置的简易算法

矩阵转置的算法研究目前主要集中在 SAR 实时成像处理, 稀疏矩阵的转置算法和多处理器并行的转置算法这些方面。接下来对不同研究方向的具有代表性的几种方法进行简要分析, 最后在一般矩阵的经典算法基础上, 提出一种多维矩阵转置的简易算法。

SAR 即星载合成孔径雷达, 是一种基于微波成像的主动式传感器, 通过发射宽带信号并应用合成孔径技术以获得二维高分辨率图像。在星载 SAR 的实时成像处理过程中, 成像算法需要多次交替对距离向和方位向的数据进行处理, 为保证数据在两个维度上的数据访问效率均衡, 防止存储器数据访问带宽受限, 需要进行存储器数据矩阵转置算法研究, 以提高转置效率, 从而提高 SAR 成像系统的成像实时性^[2]。文献[3-4]对两页式、三页式、行进列式转置算法进行了研究, 需要消耗较多的存储资源; 文献[5]提出矩阵分块线性映射法存储数据, 通过平衡两维数据的跳行次数来提高转置效率; 文献[6]提出的矩阵分块线性映射法, 以 SDRAM 作为外部高速存储器, 通过改变 SAR 数据矩阵的存储方式来提高转置效率; 文献[7]研究了多计算机上矩阵的分布以及分布式矩阵的转置, 为并行分布式转置方法提供了参考; 文献[8]给出了适用于 SMP 系统的交织存储访问矩阵转置算法, 通过优化页内数据分布来提高矩阵转置效率。

稀疏矩阵指矩阵元素中存在很多相同值的元素或零元素, 其余元素呈不规则分布。稀疏矩阵转置方法的研究主要集中在非零元素的存储方法研究。经典的存储方法是三元组线性表方式, 即每一个非零元素由一个三元组 (i, j, value) 表示, 三元组中 i, j 表示矩阵元素的行号、列号, value 表示数值, 矩阵中所有非零元素构成一个三元组线性表。文献[9]基于伪地址存储结构提出了一种伪地址存储方法, 设计一种伪地址计算公式, 将非零元素由三元组转换成二元组存储, 一定程度上减少了存储空间。

多处理器并行转置算法的研究, 主要解决海量数据以及多维矩阵的转置问题, 特点是面向不同的硬件平台展开设计与研究, 文献[10]提出了基于 OpenCL 框架的 CPU 和 GPU 并行的转置算法, 获得了比 CPU 矩阵转置串行算法更快的加速比; 文献[11]则是基于国防科技大学自主研发的一款异构

多核数字信号处理器 (DSP), 研究了不同数据位宽矩阵的并行转置算法, 转置性能和硬件的利用率都有一定提升, 为国产硬件平台的推广应用提供了支持。

综上分析, 矩阵的存储方式不同, 采用的硬件平台不同, 相对应的转置方法也不同。在 SAR 成像、数值分析领域, 矩阵转置算法通常只研究两个维度, 常采用二维数组存储, 对应于矩阵的行和列。在多维图像、信号处理领域, 多维矩阵的存储和转置处理也必不可少, 但针对多维矩阵的转置算法研究相对较少。在二维矩阵的经典算法基础上^[12], 提出一种多维矩阵转置的简易算法。下面的算法基于 Python 语言进行描述。

- (1) 定义一个函数 `matrix`, 输入参数为一个多维矩阵。
- (2) 获取矩阵的维度 `dim`, 即矩阵中包含坐标轴的数量。
- (3) 如果维度 `dim` 小于 2, 返回原始矩阵, 不进行转置操作。

- (4) 对于维度大于等于 2 的矩阵, 执行如下步骤:

① 获取原矩阵 `matrix` 的形状 `shape`, 利用矩阵转置坐标轴的对换规律确定转置后新矩阵的形状; # 如 `matrix.shape=a*b*c*d, matrix.transpose(3, 0, 1, 2).shape=d*a*b*c`;

② 根据新矩阵的形状创建一个空的矩阵 `transposed_matrix`, 并标记为待填充的转置后的矩阵;

③ 遍历原矩阵 `matrix` 中的每一个元素索引, 执行以下步骤:

步骤 1: 根据当前元素索引 `index` 和转置操作, 将 `index` 坐标轴对换, 获取转置后矩阵中的新位置 `new_index`; # 如 `matrix[a][b][c][d]=matrix.transpose(3,0,1,2)[d][a][b][c]`;

步骤 2: 将新矩阵 `new_index` 位置填充当前索引 `index` 元素的值;

- ④ 返回转置后的矩阵 `transposed_matrix`。

通过以上算法可以快速实现多维矩阵元素的转置操作。对于多维矩阵的存储方法, 适用的硬件平台以及算法效率的研究, 是以后需要加强的方向。

4 结语

矩阵转置在科学计算和图像处理的应用广泛, 文章基于 Python 的 NumPy 库, 分析了 3 种矩阵转置方法的特点, 针对 Transpose 方法提出一种坐标轴的图形表示, 可以帮助研究者更好地理解转置过程中元素的位置变换; 在二维矩阵的经典转置算法基础上, 给出了一种多维矩阵转置简易算法。在以后的研究中, 可以加强高维矩阵的存储方法以及适用的硬件平台的研究。

参考文献:

- [1] 曹富军, 袁冬芳. 矩阵转置对图形变换的几何意义 [J]. 内蒙古科技大学学报, 2021, 40(2): 112-116.

(下转第 191 页)

- [5] 王贡献,付泽,胡志辉,等.基于多分支卷积神经网络的轴承变工况故障诊断[J].噪声与振动控制,2023,43(6):135-141.
 - [6] 郝洪涛,王凯,张炳建,等.多尺度特征自适应融合的气动控制阀故障诊断[J].仪器仪表学报,2023,44(10):167-178.
 - [7] 郭宝良,赵玉秀,史丽晨,等.双轴惯性直线振动机械滚动轴承单一局部损伤故障诊断[J].振动与冲击,2023,42(24):24-31.
 - [8] LI Y H, MEN Z H, BAI X N, et al. A bearing fault diagnosis method based on M-SSCNN and M-LR attention mechanism[J/OL]. Structural health monitoring, 2024[2024-11-10].https://www.researchgate.net/publication/380202307_A_bearing_fault_diagnosis_method_based_on_M-SSCNN_and_M-LR_attention_mechanism. DOI:10.1177/14759217241244477.
 - [9] HU C Q, LI Y H, CHEN Z, et al. Research on fault diagnosis of rolling bearing based on multi-sensor bi-layer information fusion under small samples[J]. Review of scientific instruments, 2023, 94(11): 115106.
 - [10] LI G, AO J X, HU J Y, et al. Dual-source gramian angular field method and its application on fault diagnosis of drilling pump fluid end[J]. Expert systems with applications, 2024, 237: 121521.
 - [11] MEMARIAN A, DAMARLA S K, MEMARIAN A, et al. Detection of poor controller tuning with gramian angular field (GAF) and stack auto encoder (SAE)[J]. Computers and chemical engineering, 2024, 185: 108652.
 - [12] GUO Y J, LEI X J, PAN Y. An encoding-decoding framework based on CNN for circRNA-RBP binding sites prediction[J]. Chinese journal of electronics, 2024, 33(1): 256-263.
 - [13] JORGES C, BERKENBRINK C, GOTTSCHALK H. Spatial ocean wave height prediction with CNN mixed-data deep neural networks using random field simulated bathymetry[J]. Ocean engineering, 2023, 271: 113699.
 - [14] 雷亚国,韩天宇,王彪,等. XJTU-SY 滚动轴承加速寿命试验数据集解读[J]. 机械工程学报, 2019, 55(16): 1-6.
- 【作者简介】**
- 谢胜灵(1999—)男,湖南长沙人,硕士,研究方向:故障诊断、类不平衡, email: 3347313819@qq.com。
- 荆晓远(1971—),通信作者(email: jingxy_2000@126.com),男,江苏南京人,博士,教授、博士生导师,研究方向:模式识别、计算机视觉、故障诊断。
- 杜杰宾(2000—),男,广东深圳人,硕士研究生,研究方向:故障诊断、类不平衡。
- 姚永芳(1975—),女,湖南常德人,硕士,助理研究员,研究方向:模式识别、机器学习、人工智能与应用。
- 易鹏(1986—)男,湖北武汉人,博士,研究方向:计算机科学、人工智能研究。
- 王平雁(1990—)男,广东湛江人,博士,研究方向:软件安全、云数据安全、人工智能。
- (收稿日期:2025-01-09 修回日期:2025-05-16)

.....

(上接第 186 页)

- [2] 张傲,李涌睿,张至涵,等.星载 SAR 实时成像系统矩阵转置方法综述[C]//中国高科技产业化研究会智能信息处理产业化分会.第十七届全国信号和智能信息处理与应用学术会议论文集.北京:北京理工大学信息与电子学院雷达技术研究所,2023:111-235.
 - [3] 谢应科,张涛,韩承德.实时 SAR 成像系统中矩阵转置的设计和实现[J].计算机研究与发展,2003(1):6-11.
 - [4] 卢世祥,韩松,王岩飞.合成孔径雷达实时成像转置存储器的两页式结构与实现[J].电子与信息学报,2005(8):1226-1228.
 - [5] 林桐,谢宜壮,刘伟.实时 SAR 成像系统矩阵原位转置的实现[J].计算机工程,2013,39(6):319.
 - [6] 边明明,毕福昆,汪精华.实时 SAR 成像系统矩阵转置方法研究与实现[J].计算机工程与应用,2011,47(22):117-119.
 - [7] 江帆,刘光平,周志敏.多计算机上分布式矩阵转置[J].微处理机,2002(2):34-37.
 - [8] 远远,陈亮.基于 SMP 的矩阵转置算法研究与实现[J].计算机工程与设计,2016,37(10):2690-2694.
 - [9] 任志国,侯永艳.基于伪地址存储结构的稀疏矩阵快速转置算法[J].仪器仪表与自动化装置,2019(5):125-128.
 - [10] 肖汉,李彩林,李琦,等.CPU+GPU 异构并行的矩阵转置算法研究[J].东北师大学报(自然科学版),2019,51(4):70-77.
 - [11] 裴向东,王庆林,廖林玉,等.多核数字信号处理器并行矩阵转置算法优化[J].国防科技大学学报,2023,45(1):57-66.
 - [12] 刘玉坤,单冬红.矩阵转置算法的研究[J].计算机时代,2011(9):30-33.
- 【作者简介】**
- 王莉丽(1981—),女,安徽临泉人,硕士研究生,副教授,研究方向:智能信息系统。
- (收稿日期:2024-06-26 修回日期:2025-04-30)