arXiv:2601.04366v1 [cs.LG] 7 Jan 2026

用于稀疏成对比较矩阵补全的机器学习模型

Selcuk Koyuncu[a, *]，Ronak Nouri[a]，Stephen Providence[a]

[a]数学与计算机科学系，科平州立大学，美国马里兰州巴尔的摩

摘要

本文提出了一种用于稀疏成对比较矩阵（PCM）的机器学习模型，将经典的成对比较矩阵方法与基于图的学习技术相结合。通过数值结果展示了所提方法的有效性和可扩展性。

关键词： 成对比较矩阵，稀疏性，排序，图学习，图神经网络，
机器学习 2020年数学主题分类：90B50，68T05，15A18

1. 引言

令 $A=(a_{ij})_{n\times n}$ 为一个与备选项 $A_1, A_2, \ldots, A_n$相关的 n 阶方阵。若 A 满足以下条件，则称其为成对比较矩阵（*PCM*）：

1. 对所有 i, j， $a_{ij} > 0$；
2. 对所有 i， $a_{ii} = 1$；
3. 对所有 $i \neq j$，
$a_{ij} = \frac{1}{a_{ji}}$。

这里， a_{ij} 表示备选项 A_i相对于 A_j的相对偏好或重要性。如果矩阵 A 还满足以下条件，则称其为一致的

$$a_{ij} \cdot a_{jk} = a_{ik}, \quad \text{for all } i, j, k.$$

*通讯作者。电子邮箱：skoyuncu1@gmail.com（Selcuk Koyuncu），rnouri@coppin.edu（RonakNouri），sprovidence@coppin.edu（Stephen Providence）

当人类判断或观察到的交互本质上是相对的而非绝对的时，就会出现成对比较。示例包括决策分析中的偏好 elicitation、最佳–最差式提问、众包和同伴评分、体育及锦标赛结果，以及推荐系统中的隐式反馈。在这些情境下，获取所有 $\binom{n}{2}$ 比较并不可行，因此通常只能观察到其中一个较小的子集 $|\Omega| \ll n^2$。

传统的层次分析法应用通常涉及中等规模的 n，而在现代数据驱动型应用中，大规模情形自然会出现：例如，项目可能对应产品、网页或用户，此时 n可能达到10^5–10^7，观测到的比较数量则更接近 $O(n\log\, n)$，甚至达到 $O(n)$。在这些情形下，核心计算挑战在于推断缺失的比较（矩阵补全），并从稀疏、含噪且可能不一致的观测中提取可靠的全局排名，同时保持成对比较矩阵的互反结构并促进乘法一致性。这促使人们开发可扩展的学习算法，使其直接在稀疏图上运行，而非在稠密的 $n \times n$ 矩阵上运行。

成对比较矩阵（PCM）由Saaty于20世纪70年代末提出，是层次分析法（AHP）的基础 [3]。成对比较矩阵对备选项之间的相对偏好进行编码；当存在权重向量 w ，使得 $a_{ij} = w_i/w_j$ 对所有i，j均成立时，该矩阵具有一致性。实际中，人类判断很少具有一致性，因此人们提出了多种指标来衡量不一致性，包括Saaty的一致性比率和Koczkodaj三元组指标 [4]。从成对比较矩阵中推导优先级的方法包括特征向量法、几何平均法（等价于求解对数最小二乘问题 [5]），以及与Bradley–Terry–Luce和Thurstone模型相关的统计方法。

当成对比较矩阵不一致时，一个自然的任务是寻找与之最近的一致矩阵；一般而言，这是一个NP难问题 [6]。为应对这一挑战，研究者提出了凸松弛和最小二乘形式等近似方法。近年来，研究者还发现了成对比较矩阵与基于图的排名方法之间的联系。Rank Centrality [7] 等谱算法运用马尔可夫链技术，在具有可证明保证的前提下恢复排序；学习理论的研究结果则表明，通常只需进行 $O(n\log\, n)$ 次成对比较，就能获得准确的排序。

除经典决策分析外，成对比较矩阵已应用于群体决策、推荐系统、体育分析和众包，并扩展至区间和模糊情形。尽管取得了这些进展，大多数

现有理论仍假设成对比较矩阵是稠密的。实际上，当备选项数量较多时，收集所有 $\binom{n}{2}$ 比较并不可行，因此成对比较矩阵通常高度稀疏。因此，如何补全稀疏成对比较矩阵、控制不一致性并提取稳健的全局排名，已成为一个新兴且重要的研究方向，处于矩阵理论、优化和机器学习的交叉领域。

虽然一致性指标（CI）和一致性比率（CR）等一致性度量能提供有用的诊断信息，但它们并未直接解决稀疏或高度不一致的成对比较矩阵的补全问题。在本研究中，我们引入了一种机器学习模型，可在保持乘法一致性的同时重建缺失的比较。

我们给出以下简单示例，以提醒读者这一基本思想。

考虑三个备选项 A_1、A_2、A_3。假设它们的真实权重为w =(2，1，0.5)。那么成对比较矩阵为

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 1/2 & 1 & 2 \\ 1/4 & 1/2 & 1 \end{bmatrix}.$$

该矩阵是一致的，因为 $a_{ij} \cdot a_{jk} = a_{ik}$ 对所有 i、j、k均成立。例如，$a_{12} \cdot a_{23} = 2 \cdot 2 = 4 = a_{13}$。

现在，假设给定以下判断：

$$B = \begin{bmatrix} 1 & 3 & 4 \\ 1/3 & 1 & 2 \\ 1/4 & 1/2 & 1 \end{bmatrix}.$$

该矩阵不一致，因为 $a_{12} \cdot a_{23} = 3 \cdot 2 = 6 \neq a_{13} = 4$。这种不一致在现实世界的决策中很常见，因为人类判断并非完全具有传递性。

考虑

$$B = \begin{bmatrix} 1 & 3 & 4 \\ \frac{1}{3} & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & 1 \end{bmatrix}.$$

步骤 *1*：优先向量.. 计算 B的主右特征向量 w （归一化至总和为 1）。数值上，

$$w \approx (0.6250,\ 0.2385,\ 0.1365)^T.$$

步骤 *2*：最大特征值.. 令 $\lambda_{\max}$ 为 B的最大特征值。数值上，

$$\lambda_{\max} \approx 3.0183.$$

步骤 *3*：一致性指标（*CI*）.. 对于 $n = 3$,

$$CI \;=\; \frac{\lambda_{\max} - n}{n-1} \;=\; \frac{3.0183 - 3}{2} \;\approx\; 0.00915.$$

St骤 *4*：一致性比率（*CR*）。使用 Saaty随机指标 $RI(3) = 0.58$,

$$CR \;=\; \frac{CI}{RI} \;=\; \frac{0.00915}{0.58} \;\approx\; 0.0158.$$

由于 $CR \approx 0.016 < 0.10$，矩阵 B 并非完全一致，例如 $a_{12}a_{23} = 6 \neq a_{13} = 4$，但根据 Saaty准则，其一致性可接受 。

本研究旨在补全稀疏互反成对比较矩阵并恢复全局排名，同时做到：(i) 仅利用观测到的比较，(ii) 可扩展至大型图，(iii) 通过在对数空间中采用基于三角形的正则化，明确促进乘法一致性。我们的方法将经典成对比较矩阵思想（对数比建模、互反性和传递性）与基于图的表示学习相结合，从而得到一种实用的补全方法；在合成基准测试中的准确度可与标准基线方法相媲美，并且在稀疏大规模场景下仍具备计算可行性。

稀疏成对比较矩阵（PCM）是一个只部分填充、含有许多缺失项的矩阵，因此重建完整的比较结构并提取一致性排名颇具挑战。本文提出一种融合式机器学习模型，用于补全任意规模成对比较矩阵中的缺失项并确保其一致性。例如，矩阵中的“? ”表示缺失项，我们的方法提供了一套系统的补全流程，同时保持两两一致性：

$$A = \begin{bmatrix} 1 & 3 & ? & ? & ? \\ \frac{1}{3} & 1 & 5 & ? & ? \\ ? & \frac{1}{5} & 1 & 2 & ? \\ ? & ? & \frac{1}{2} & 1 & 4 \\ ? & ? & ? & \frac{1}{4} & 1 \end{bmatrix}.$$

大量文献研究了不完整的成对比较矩阵，以及推导可靠优先级所需的比较数量。尤其是，近期研究探讨了在何种情况下可以容忍不完整性，而不会导致过度不一致或不稳定。Agóston 和 Csató [12] 研究了不完整的成对比较矩阵的不一致性阈值，并给出了不完整判断仍具有意义的条件。Brunelli [13] 讨论了仅使用 $(n-1)$ 次比较的局限性，并回顾了为何可能需要更多比较以增强稳健性。这些研究强调，稀疏性不仅是计算问题，也是统计和决策理论问题：不同的观测模式可能导致截然不同的不确定性和不一致性表现。

稀疏成对比较矩阵补全与机器学习中广泛使用的矩阵补全和低秩分解方法密切相关，尤其是在推荐系统中。经典方法通过低秩分解近似部分观测矩阵，并通过正则化优化来估计潜在因子。在成对比较场景中，诸如 $a_{ij} \approx \exp(x_i - x_j)$这样的对数比参数化，可以视为对数空间中的一种结构化（秩二）分解。我们的基于图的模型可解释为一种非线性扩展，其中节点嵌入取代标量得分，并通过在观测到的比较图上进行消息传递来不断优化。

另一个相关研究方向是基于部分比较结果进行拟合的概率排序与选择模型（Bradley–Terry模型、Plackett–Luce模型及相关模型族）。这类模型提供了一种原则性方法，可将稀疏结果汇总为全局得分，并推导出一致的偏好概率。例如，Cheng等人 [14]讨论了使用阈值化概率模型处理部分弃权情况下的标签排序，说明学习全局模型并进行阈值化如何得出连贯的部分偏好。我们的框架与这一观点兼容：BTL模式损失用于拟合观测结果，而一致性正则项则促进所推导偏好结构的传递性。

本文结构如下：第 1 节和第 2 节分别概述两种广为人知的成对比较矩阵基线方法。第 4 节介绍所提出的机器学习模型，并给出若干说明性示例。第 5 节报告数值实验结果，展示该模型的性能。最后，第 6 节讨论处理大规模图时的可扩展性和效率，并给出其他数值结果。

现在介绍两种著名的模型。

2. Bradley–Terry–Luce（BTL）模型

Bradley–Terry–Luce（BTL）模型 [1, 2] 是一种用于成对比较的经典概率框架。假设有 n 个备选项$A_1, \ldots, A_n$。每个备选项 A_i 都被赋予一个潜在得分 $x_i \in \mathbb{R}$。将 A_i 优于 A_j 的概率建模为

$$\Pr(A_i \succ A_j) = \sigma(x_i - x_j) = \frac{e^{x_i}}{e^{x_i} + e^{x_j}}, \quad i \neq j,$$

其中， $\sigma(\cdot)$ 表示逻辑 Sigmoid 函数。

最大似然估计

给定成对比较的观测结果，通过最大化对数似然来估计潜在得分 x =(x1、……、 xn)

$$\ell(x) = \sum_{i \neq j} \Big[c_{ij} \, \log \sigma(x_i - x_j) + c_{ji} \, \log \sigma(x_j - x_i) \Big],$$

其中， c_{ij} 是 A_i 击败 A_j的次数。该优化问题是凸的（除一个加法常数外，因为 x 仅在整体平移意义下可识别）。通常，通过施加约束 $\sum_i x_i = 0$ 来固定规范。

3. 对数最小二乘（LLS）方法

虽然布拉德利–特里–卢斯模型（BTL）是概率模型，但对数最小二乘（LLS）方法是一种用于补全和近似成对比较矩阵（PCMs）的确定性方法。设 A =(a_{ij}) 是一个包含若干观测条目的成对比较矩阵，其中 $a_{ij} > 0$ 且 $a_{ij} = 1/a_{ji}$。LLS 方法假设存在潜在得分 x =(x1、……、 xn)，使得

$$a_{ij} \approx e^{x_i - x_j}.$$

通过求解以下凸优化问题来估计潜在得分

$$\min_{x \in \mathbb{R}^n} \sum_{(i,j) \in \Omega} \big(x_i - x_j - \log a_{ij} \big)^2,$$

其中 Ω 表示观测条目的集合。这归结为求解一个涉及由 Ω 诱导的图拉普拉斯矩阵的线性方程组。

4. 用于稀疏成对比较矩阵补全的机器学习模型

我们将稀疏成对比较矩阵（PCM）建模为图 $G = (V,\ E)$，其中包含 $|V| = n$ 个项目，边 $E \subseteq V \times V$ 上存在观测到的比较。令 $\Omega \subseteq \{(i,\ j) : i \neq j\}$ 表示观测到的有向比较；每个 $(i,\ j) \in \Omega$ 携带以下两种结果之一：(i) 二元结果 $y_{ij} \in \{0, 1\}$（*BTL* 模式），或 (ii) 基数比值 $a_{ij} > 0$（*LLS* 模式）。对于基数数据，我们在对数空间中使用目标 $t_{ij} = \log a_{ij}$。

我们首先介绍潜在嵌入和边预测器，每个项目 i 都有一个可学习嵌入 $h_i \in \mathbb{R}^d$。浅层消息传递网络通过聚合邻居来更新嵌入：

$$h_i^{(0)} \in \mathbb{R}^d, \qquad h_i^{(\ell+1)} = \phi\Big(W_1\, h_i^{(\ell)} + \sum_{j \in \mathcal{N}(i)} W_2\, h_j^{(\ell)}\Big), \quad \ell = 0, \ldots, L-1,$$

其中包含非线性函数 ϕ 和学习得到的权重 W_1，W_2。给定最终嵌入 $h_i = h_i^{(L)}$，我们通过作用于差值的线性边预测头来预测（对数）比较值：

$$\widehat{t}_{ij} = v^\top (h_i - h_j) \quad \Longrightarrow \quad \widehat{a}_{ij} = \exp(\widehat{t}_{ij}), \qquad \widehat{p}_{ij} = \sigma\big(v^\top (h_i - h_j)\big),$$

其中 $\widehat{p}_{ij}$ 是 BTL 模式下使用的概率 $\Pr(i \succ j)$，$v \in \mathbb{R}^d$ 是维度为 d的可学习权重向量，σ 是逻辑 Sigmoid 函数。

对于训练目标，我们拟合观测边，并强制满足全局乘法一致性：

$$\mathcal{L}_{\text{data}} = \begin{cases} -\sum_{(i,j)\in\Omega} \big[y_{ij} \log \widehat{p}_{ij} + (1 - y_{ij}) \log(1 - \widehat{p}_{ij})\big], & \text{BTL mode,} \\ \sum_{(i,j)\in\Omega} \big(\widehat{t}_{ij} - t_{ij}\big)^2, & \text{LLS mode,} \end{cases}$$

$$\mathcal{L}_{\triangle} = \frac{1}{|\mathcal{T}|} \sum_{(i,j,k)\in\mathcal{T}} \Big|\widehat{t}_{ij} + \widehat{t}_{jk} - \widehat{t}_{ik}\Big|, \qquad \mathcal{L}_{\text{reg}} = \|W_1\|_F^2 + \|W_2\|_F^2 + \|v\|_2^2.$$

这里，$\mathcal{T}$是节点三元组的随机集合（三角形采样），$\widehat{t}_{ij} = \log \widehat{a}_{ij}$在两种模式下均适用。总损失为

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_{\triangle}\, \mathcal{L}_{\triangle} + \lambda_{\text{reg}}\, \mathcal{L}_{\text{reg}}.$$

最小化 $\mathcal{L}_{\triangle}$ 有助于在对数空间中实现加法传递性，即比率的乘法一致性。

下一步，我们可以进行预测与投影。训练后，我们通过 $\widehat{a}_{ij} = \exp(v^\top(h_i - h_j))$（LLS 模式）预测缺失项，或通过优势比 $\widehat{a}_{ij} = \widehat{p}_{ij}/(1-\widehat{p}_{ij})$（BTL 模式）进行预测。然后，我们将其投影 为严格互反的成对比较矩阵：设置 $\widehat{a}_{ii} = 1$ ，并将每一对 $(\widehat{a}_{ij}, \widehat{a}_{ji})$ 替换为通过几何均值投影得到的最近互反数对：

$$\widetilde{a}_{ij} \leftarrow \sqrt{\widehat{a}_{ij} \cdot \frac{1}{\widehat{a}_{ji}}}, \qquad \widetilde{a}_{ji} \leftarrow \frac{1}{\widetilde{a}_{ij}}.$$

最后，我们来看计算与可扩展性。每个训练轮次的数据损失计算成本为 $O(|\Omega|d)$，三角形采样成本为 $O(|\mathcal{T}|d)$（选择 $|\mathcal{T}| \ll n^3$）。对于非常大的 n，每个批次都对子采样边，并且仅在最后计算稠密全配对矩阵。

现在，我们通过一个示例来说明该方法的细节。

示例 4.1. 我们通过一个包含 $n = 4$ 个项目 A_1、A_2、A_3、A_4 的简单两两比较图，给出潜在嵌入和边预测机制。假设我们观测到以下比较

$$a_{12} = 3, \qquad a_{23} = 5, \qquad a_{34} = 2,$$

在这种情况下，成对比较矩阵为

$$A = \begin{bmatrix} 1 & 3 & ? & ? \\ \frac{1}{3} & 1 & 5 & ? \\ ? & \frac{1}{5} & 1 & 2 \\ ? & ? & \frac{1}{2} & 1 \end{bmatrix}$$

观测边集为 $\Omega = \{(1,2)、(2,3)、(3,4)\}$。目标是为缺失项赋值，使补全后的矩阵满足两两一致性。

步骤 *1*：启动时，为每个项目 i 分配一个初始嵌入向量 $h_i^{(0)} \in \mathbb{R}^2$。例如，我们设定

$$h_1^{(0)} = (0.2, -0.1), \quad h_2^{(0)} = (-0.3, 0.4), \quad h_3^{(0)} = (0.1, 0.0), \quad h_4^{(0)} = (-0.2, -0.5).$$

步骤 *2*：进行消息传递更新时，我们通过以下方式更新每个嵌入

$$h_i^{(1)} = \phi\Big(W_1 h_i^{(0)} + \sum_{j\in\mathcal{N}(i)} W_2 h_j^{(0)}\Big),$$

其中 $W_1 = I$, $W_2 = 0.5I$。

对于节点 A_2:

$$h_2^{(1)} = \phi\big(h_2^{(0)} + 0.5(h_1^{(0)} + h_3^{(0)})\big).$$

从数值上看，

$$\begin{aligned} h_2^{(1)} &= \phi\big((-0.3, 0.4) + 0.5((0.2, -0.1) + (0.1, 0.0))\big) \\ &= \phi\big((-0.3, 0.4) + (0.15, -0.05)\big) \\ &= \phi(-0.15,\ 0.35). \end{aligned}$$

步骤 *3*：对 $\phi(x) = \max(0, x)$ 逐分量应用 ***ReLU***。

$$h_2^{(1)} = \phi(-0.15,\ 0.35) = (0,\ 0.35).$$

同理：

$h_1^{(1)} = \phi\big((0.2, -0.1) + 0.5(-0.3, 0.4)\big) = \phi(0.05,\ 0.1) = (0.05,\ 0.1),$

$h_3^{(1)} = \phi\big((0.1, 0.0) + 0.5((-0.3, 0.4) + (-0.2, -0.5))\big) = \phi((-0.35, -0.05)) = (0,\ 0),$

$h_4^{(1)} = \phi\big((-0.2, -0.5) + 0.5(0.1, 0.0)\big) = \phi((-0.15, -0.5)) = (0,\ 0).$

因此，更新后的嵌入为

$$h_1^{(1)} = (0.05,\ 0.1), \quad h_2^{(1)} = (0,\ 0.35), \quad h_3^{(1)} = (0,\ 0), \quad h_4^{(1)} = (0,\ 0).$$

步骤 *4*：对于边预测，取 $v = (1,1)^\top$，然后

$$\widehat{t}_{ij} = v^\top (h_i - h_j), \qquad \widehat{a}_{ij} = \exp(\widehat{t}_{ij}).$$

对于边 (1,2)：

$h_1 - h_2 = (0.05, 0.1) - (0, 0.35) = (0.05, -0.25), \quad \widehat{t}_{12} = 1\cdot 0.05 + 1\cdot(-0.25) = -0.20.$

Thus

$$\widehat{a}_{12} = \exp(-0.20) \approx 0.82.$$

同理，(2,3)：

$$h_2 - h_3 = (0, 0.35) - (0, 0) = (0, 0.35), \quad \widehat{t}_{23} = 0.35, \quad \widehat{a}_{23} = e^{0.35} \approx 1.42.$$

步骤 *5*：现在，我们使用最终嵌入 h_i和边预测头 v补全所有缺失项，预测每一对(i, j) ，方法是

$$\widehat{t}_{ij} \;=\; v^{\top}(h_i - h_j), \qquad \widehat{a}_{ij} \;=\; \exp(\widehat{t}_{ij}), \quad i \neq j,$$

$\widehat{\ }$并设置 $\widehat{a}_{ii} \;=\; 1$。这会得到一个稠密矩阵 $A \;=\; (\widehat{a}_{ij})$，但它可能并非完全互反，因为两个方向的预测 $\widehat{a}_{ij}$ 和 $\widehat{a}_{ji}$ 来自独立的前向传播。

步骤 *6*：将每个无序对 $\{i, j\}$ 通过几何均值投影投影到最近的互反对：

$$\widetilde{a}_{ij} \;\leftarrow\; \sqrt{\widehat{a}_{ij} \cdot \frac{1}{\widehat{a}_{ji}}}, \qquad \widetilde{a}_{ji} \;\leftarrow\; \frac{1}{\widetilde{a}_{ij}}, \qquad \widetilde{a}_{ii} = 1.$$

$\widetilde{\ }$矩阵 $A \;=\!(\widetilde{a}_{ij})$ 现在是一个有效的成对比较矩阵，即为正、具有单位对角线且$\widetilde{\ }$互反。如果训练时使用了三角一致性惩罚， A 也将具有近似一致性。

步骤 *7*：注意，以下是通过 *LLS* 得到的精确一致补全。如果需要精确的乘法一致补全，则在观测边 $\Omega \;=\{(1,2),(2,3),(3,4)\}$上求解对数最小二乘（*LLS*）问题：

$$\min_{x \in \mathbb{R}^4} \sum_{(i,j)\in\Omega} \big(x_i - x_j - \log a_{ij}\big)^2.$$

对于这条链，最优差值是精确的：

$$x_1 - x_2 = \log 3, \quad x_2 - x_3 = \log 5, \quad x_3 - x_4 = \log 2,$$

因此，一致的补全结果为

$$\widehat{A} \;=\; \exp(x_i - x_j) \;=\; \begin{bmatrix} 1 & 3 & 15 & 30 \\ \frac{1}{3} & 1 & 5 & 10 \\ \frac{1}{15} & \frac{1}{5} & 1 & 2 \\ \frac{1}{30} & \frac{1}{10} & \frac{1}{2} & 1 \end{bmatrix}.$$

示例 4.2。 我们有五个备选项 $A_1, \ldots, A_5$，其观测到的基数比率

$$a_{12} = 3, \qquad a_{23} = 5, \qquad a_{34} = 2, \qquad a_{45} = 4,$$

其余所有条目均缺失，这个成对比较矩阵

$$A = \begin{bmatrix} 1 & 3 & ? & ? & ? \\ \frac{1}{3} & 1 & 5 & ? & ? \\ ? & \frac{1}{5} & 1 & 2 & ? \\ ? & ? & \frac{1}{2} & 1 & 4 \\ ? & ? & ? & \frac{1}{4} & 1 \end{bmatrix}.$$

我们在对数空间中处理 $x \in \mathbb{R}^5$ 并对 $a_{ij} \approx e^{x_i - x_j}$ 建模。

(A) LLS：我们先建立正规方程。。令 $\Omega = \{(1,2), (2,3), (3, 4), (4, 5)\}$并令 $y_{ij} = \log a_{ij}$。对于每个 $(i, j) \in \Omega$，将其加入拉普拉斯矩阵 L 和向量 b：

$$L_{ii}{+}{=}1, \quad L_{jj}{+}{=}1, \quad L_{ij}{-}{=}1, \quad L_{ji}{-}{=}1, \qquad b_i{+}{=}y_{ij}, \quad b_j{-}{=}y_{ij}.$$

With

$$y_{12} = \text{对数}\, 3 \approx 1.0986, \quad y_{23} = \text{对数}\, 5 \approx 1.6094, \quad y_{34} = \text{对数}\, 2 \approx 0.6931, \quad y_{45} = \text{对数}\, 4 \approx 1.3863,$$

我们得到

$$L = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \qquad b = \begin{bmatrix} \log 3 \\ -\log 3 + \log 5 \\ -\log 5 + \log 2 \\ -\log 2 + \log 4 \\ -\log 4 \end{bmatrix} = \begin{bmatrix} 1.0986 \\ 0.5108 \\ -0.9163 \\ 0.6932 \\ -1.3863 \end{bmatrix}.$$

(B) 求解 x。。这些方程编码了沿链的精确差值：

$$x_1 - x_2 = \log 3, \quad x_2 - x_3 = \log 5, \quad x_3 - x_4 = \log 2, \quad x_4 - x_5 = \log 4.$$

令 $x_5 = t$。那么

$$x_4 = t + \log 4, \quad x_3 = t + \log 8, \quad x_2 = t + \log 40, \quad x_1 = t + \log 120.$$

通过零均值固定规范：$x_1 + x_2 + x_3 + x_4 + x_5 = 0$ 得到

$$5t + \log(120 \cdot 40 \cdot 8 \cdot 4 \cdot 1) = 0 \quad \Rightarrow \quad t = -\tfrac{1}{5} \log(153{,}600).$$

数值上，$\log(153{,}600) \approx 11.9400$，因此 $t \approx -2.3880$。因此

$$\begin{aligned} x_1 &= t + \log 120 \approx 2.4056, \\ x_2 &= t + \log 40 \approx 1.2924, \\ x_3 &= t + \log 8 \approx -0.2072, \\ x_4 &= t + \log 4 \approx -1.0018, \\ x_5 &= t \approx -2.3880. \end{aligned}$$

(C) 接下来，我们补全成对比较矩阵。设 $\hat{a}_{ij} = e^{x_i - x_j}$、$\hat{a}_{ji} = 1/\hat{a}_{ij}$ 和 $\hat{a}_{ii} = 1$。补全后的一致成对比较矩阵为

$$\widehat{A} = \begin{bmatrix} 1 & 3 & 15 & 30 & 120 \\ \frac{1}{3} & 1 & 5 & 10 & 40 \\ \frac{1}{15} & \frac{1}{5} & 1 & 2 & 8 \\ \frac{1}{30} & \frac{1}{10} & \frac{1}{2} & 1 & 4 \\ \frac{1}{120} & \frac{1}{40} & \frac{1}{8} & \frac{1}{4} & 1 \end{bmatrix}.$$

(D) 例如，要检查一致性，

$$\log \hat{a}_{12} + \log \hat{a}_{23} - \log \hat{a}_{13} = \log 3 + \log 5 - \log 15 = 0.$$

$\widehat{\ }$所有其他三角形也满足类似的等式，因此 A 恰好是乘法一致的。

(E)。三角损失

$$\mathcal{L}_{\triangle} = \tfrac{1}{|\mathcal{T}|} \sum_{(i,j,k)} \left| \log \hat{a}_{ij} + \log \hat{a}_{jk} - \log \hat{a}_{ik} \right|$$

为 0（每一项都为零）。

(F)。布拉德利–特里–卢斯模型使用 $\sigma(u) = 1/(1+e^{-u})$对 $\Pr(i \succ j) = \sigma(x_i - x_j)$ 建模。使用相同的分数差：

$$\begin{aligned} \Pr(1 \succ 2) &= \sigma(\log 3) = 0.75, \\ \Pr(2 \succ 3) &= \sigma(\log 5) \approx 0.8333, \\ \Pr(3 \succ 4) &= \sigma(\log 2) \approx 0.6667, \\ \Pr(4 \succ 5) &= \sigma(\log 4) = 0.8. \end{aligned}$$

未观测的配对关系可通过传递性推断，例如：

$$\Pr(1 \succ 5) = \sigma(\log 120) = \tfrac{120}{121} \approx 0.9917.$$

5. 数值结果

我们的数值实验直接比较了经典的对数最小二乘（LLS）方法与我们提出的机器学习（ML）模型，测试对象是规模 n 和边密度p各异的合成稀疏成对比较矩阵。表 1 显示，两种方法在留出对数比的均方根误差（RMSE）以及排序恢复的 Kendall's τ 方面，准确度非常相近。正如预期，边密度越高，准确度越好：当仅观测到 1% 的比较（$p = 0.01$）时，误差相对较高；但到$p = 0.05$ 时，两种方法都能恢复几乎完美的排序（$\tau \approx 0.98$）。这证实，只要有一定比例的比较数据，稀疏成对比较矩阵就能得到可靠的补全。

图 1 和图 2 进一步展示了这些趋势。在准确性方面，机器学习方法在所有设置下都与 LLS 表现接近，说明它既能作为有竞争力的替代方案，也具有更强的建模灵活性。在效率方面，图 2 展示了明显的权衡：LLS 通过求解线性方程组，速度快了几个数量级；而最大似然则需要迭代训练。尽管如此，最大似然仍具备良好的可扩展性，其运行时间随观测边数量线性增长，因此适用于中大型问题。总体而言，这些结果表明，尽管 LLS 仍是效率最高的基线方法，我们的最大似然框架实现了相当的性能，并为更丰富的扩展提供了可能。

表 1 对比了 LLS 和最大似然方法在合成稀疏成对比较矩阵上的表现。我们报告实际运行时间（秒）、留出对数比的均方根误差，以及 Kendall's τ 排序相关系数。

表 1:

n	p 边	LLS 时间	LLS均方根误差	LLS τ	最大似然时间	最大似然 均方根误差	ML τ
200	0.010000	203	0.007000	0.866000	0.866000	14.231000	0.868000	0.614000
200	0.020000	426	0.011000	0.283000	0.903000	15.631000	0.285000	0.901000
200	0.050000	980	0.012000	0.162000	0.958000	17.536000	0.162000	0.958000
400	0.010000	821	0.041000	0.361000	0.886000	22.437000	0.363000	0.886000
400	0.020000	1570	0.066000	0.184000	0.951000	18.862000	0.184000	0.951000
400	0.050000	3946	0.079000	0.167000	0.976000	19.484000	0.167000	0.976000
800	0.010000	3227	0.260000	0.194000	0.955000	24.839000	0.194000	0.955000
800	0.020000	6544	0.217000	0.165000	0.972000	20.915000	0.165000	0.972000
800	0.050000	16050	0.263000	0.154000	0.984000	22.408000	0.154000	0.984000

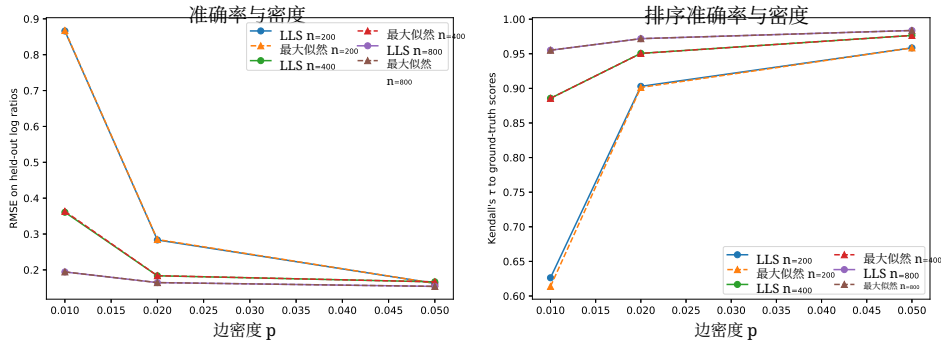


图 1：（左）留出对数比的均方根误差与边密度的关系 p。（右）肯德尔 τ 秩相关系数与边密度的关系。

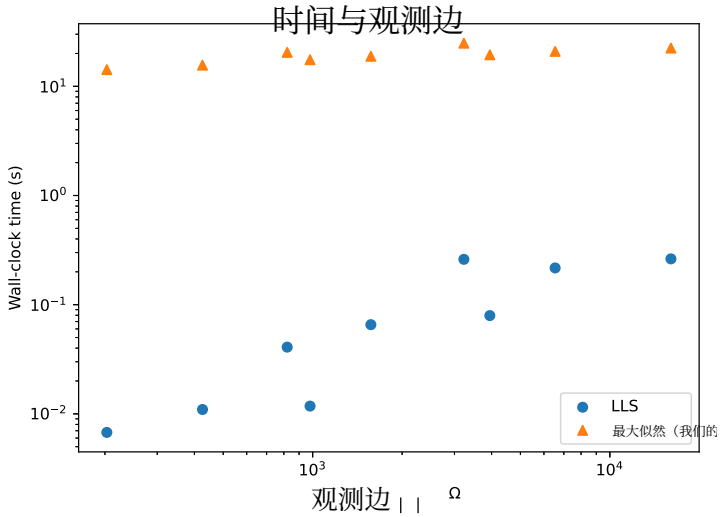


图 2：合成成对比较矩阵上，实际运行时间与观测边数 $|\Omega|$ 的关系。LLS 的速度显著快于最大似然，但两者的增长都大致呈线性。

6. 扩展与相关方法

本节在第 4 节介绍的稀疏成对比较矩阵补全核心机器学习模型基础上进行扩展，重点提升模型的可扩展性和效率，以处理大规模图。我们通过解决计算瓶颈，同时保留模型的理论基础（例如用于确保乘法一致性的三角损失），力求弥合数学优雅性与实际部署之间的差距。此外，我们将本研究置于更广泛的基于图的排名方法领域中，着重阐明与先前方法的联系和区别。

6.1. 扩展的动机

如第 4 节所示，原始模型在中等规模的合成数据集上表现有效，但在现实应用中面临可扩展性挑战；在推荐系统或社交网络排序等场景中，备选项 n 的数量可能达到数百万。每轮训练成本主要受稠密全对补全影响，复杂度为 $O(n^2)$，因而高得难以承受；而在稀疏图中，大多数三元组彼此不连通，因此用于 $\mathcal{L}_{\triangle}$ 的随机三角形采样效率低下。

受处理含有 $|\Omega| \ll n^2$ 的稀疏成对比较矩阵的需求启发，我们提出了利用稀疏运算和图感知采样的改进方案，将复杂度降至 $O(|\Omega|\log\ n)$，同时不牺牲准确性或一致性。正如引言中所讨论的，这些扩展使该模型适用于决策科学与大规模机器学习交叉领域的新兴应用。

6.2. 与先前工作的关系

我们提出的模型受到基于图的排名方法的启发，这类方法处理成对比较；但本模型的独特之处在于，着重通过三元组损失显式施加一致性约束，以补全稀疏成对比较矩阵。作为比较，我们重点介绍三种代表性方法：Rank Centrality [7], Serial Rank [8], 和 GNN Rank [9]。Rank Centrality [7] 将成对比较建模为马尔可夫链，其中转移概率反映偏好强度，并根据稳态分布得出排名。该方法在 $O(n\log\ n)$ 条比较的条件下，为准确排名提供了可证明的保证。这些比较方法假设图是完整或近乎完整的，并且除了概率推断之外，缺乏处理稀疏性或保证传递性的机制

。Serial Rank [8] 将排名问题重新表述为序列化问题，通过匹配一致性构建相似度矩阵，并利用其拉普拉斯矩阵的费德勒向量提取排名。该方法具有抗噪性，且能处理不完整数据，并为采样数据提供扰动界；但它侧重于谱松弛，而非可学习嵌入，因此限制了其与机器学习流程集成时的灵活性。GNN Rank [9]采用有向图神经网络从成对比较中学习嵌入，随后通过近端优化近似费德勒向量，从而在 NCAA 体育数据等基准测试上实现最先进的排名逆序数最小化。虽然该方法可扩展至中等规模的图（$n \sim 350$），但它优先考虑排序而非矩阵补全，也未像我们的方法 $\mathcal{L}_{\triangle}$那样引入一致性损失。

表 2：经典方法与基于图神经网络的排序方法在与我们的方法最相关的几个方面存在关键差异：稀疏性处理、一致性约束、可扩展性和可学习性。

方法	稀疏性处理	一致性 约束 ment	可扩展性重点	可学习的嵌入-向量
秩中心-性 [7]	处理稀疏比较-通过概率估计进行比较—；可证明的恢复使用 $O(n$对数 $n)$	通过平稳分布隐式实现—ary 分布 的马尔可夫链	中等 （迭代式特征向量计算）	No
Serial Rank [8]	对……具有鲁棒性 缺失/噪声条目 使用 基于相似度的序列排序	通过谱一致性	高（单个特征向量求解）	No
GNN Rank [9] 学习 from 有向通过消息通过		通过近端细化隐式实现（费德勒向量近似）	中等（$n \sim 300$–500）	Yes
Ours	稀疏 张量 迷你-批处理 and 图子采样	通过三角形或显式实现循环一致性 loss ($\mathcal{L}_{\triangle}$)	High $O(\|\Omega\|)$（迷你批次大小）	Yes

这些方法具有互补性；例如，我们的模型可以采用 Serial Rank 的序列排序作为初始化，或采用 GNN Rank 的近端步骤进行补全后排序。未来的工作可以探索混合式集成方案。

6.3. 可扩展性与效率改进

主要问题在于，原始框架每个训练轮次的复杂度为 $O(|\Omega|d+|\mathcal{T}|d+n^2)$，其中 $O(n^2)$ 稠密补全步骤会成为大型 n 的瓶颈，即使 $|\Omega|$ 是稀疏的，例如根据学习理论界为 $O(n\log n)$ 时也是如此 [7]）。在稀疏图中，随机三角形采样会降低效率，因为不连通三元组产生的梯度为零。

大多数操作都可以通过稀疏表示限制在观测边上，并且可以从连通结构中高效采样具有信息量的三角形（楔形），从而保留传递性信号。

我们的贡献在于采用基于小批量子图诱导和稀疏张量运算的高效训练流程，将每轮训练的复杂度降至 $O(|\Omega|\log\, n)$。这种改进遵循 GraphSAGE [10] 和 Cluster-GCN [11]提出的可扩展图神经网络范式。

6.3.1. 采用稀疏消息传递的小批量训练

为消除稠密矩阵，我们使用稀疏边索引张量 $E \in \mathbb{Z}^{2\times|\Omega|}$表示图，其取值为 $a \in \mathbb{R}^{|\Omega|}$。消息传递更新变为：

$$H^{(\ell+1)} = \phi\big(W_1 H^{(\ell)} + \textsc{SpMM}(A_{\text{sparse}},\, W_2 H^{(\ell)})\big),$$

其中 A_{sparse} 是稀疏邻接张量，稀疏矩阵乘法的开销为 $O(|\Omega|d)$。

对于小批量处理，对边 $\Omega_b \subset \Omega$ ($|\Omega_b| = B_e$)进行子采样，通过 $\log n$ 跳随机游走诱导子图G_b，并在 Ω_b上计算 $\mathcal{L}^b_{\text{data}}$。直观而言，子图诱导能够捕获用于嵌入更新的局部上下文，而全局同步则确保收敛。对于 $\mathcal{L}_\triangle$，楔形采样从一条边 $(i,\ j)$ 开始，并从 j 的出邻居中采样 k ——优先选择连通三元组，从而提供非零梯度，并比随机选择更高效地近似传递性。

6.3.2. 复杂度分析

我们将效率提升形式化如下：

命题 6.1. 在平均节点度有界于常数的稀疏连通性条件下，并假设小批量大小为 $B_e, B_t = O(|\Omega|/\log\, n)$，所提出的训练过程每个训练轮次的期望计算成本为$O(|\Omega|\log\, n)$。

证明。令 $G = (V, E)$ 表示具有 $|V| = n$ 和 $|E| = |\Omega|$ 的两两比较图。对于每个训练轮次，训练由两个主要部分组成：

对于稀疏消息传递：在每一层中，消息传播仅限于与边的迷你批次 $\Omega_b \subseteq \Omega$ 对应的诱导子图，且其满足 $|\Omega_b| = B_e$。每个子图包含所有在 $\log\, n$ 跳范围内可达的节点；在有界度稀疏性下，预期邻域大小为 $O(B_e \log n)$。因此，稀疏

矩阵乘法（稀疏矩阵乘法）的每层复杂度为

$$O(B_e d \log n),$$

每个训练轮次处理 $O(|\Omega|/B_e)$ 个批次，因此总成本为

$$O(|\Omega| d \log n).$$

楔形（三角形）采样：为施加三角一致性损失$\mathcal{L}_{\triangle}$，楔形采样会选取 $\mathcal{T}_b$ 个连通三元组的子集，其中每个楔形（i，j，k）至少共享一条观测边。由于信息丰富的楔形的期望数量随观测边数量线性增长，且每个楔形只需进行常数时间更新，因此这部分的成本为

$$O(B_t d),$$

其中 $B_t = O(|\Omega|/\log\ n)$ 由于邻域探索深度而额外产生一个对数因子。

合并这两项，可得每个训练轮次的期望复杂度

$$O(B_e d \log n + B_t d) = O(|\Omega| d \log n).$$

相比之下，原始模型中的稠密补全步骤的复杂度随 $O(n^2)$ 增长，对于较大的 n 难以处理。因此，在 $|\Omega| = O(n \log\ n)$ 足以实现准确的排序恢复 [7]，的稀疏情形下，所提出的方法具有近线性的可扩展性。这些界与可扩展图神经网络架构的实证结果一致 [10, 11]，，其中计算量和内存均按 $O(|\Omega|)$ 而非 $O(n^2)$ 增长。 □

6.3.3. 实现细节

该模型采用 PyTorch 实现，使用 torch.sparse（COO 布局）进行稀疏矩阵乘法（SpMM）。该实现支持 DGL 和 PyG 图接口，可无缝处理数据。超参数（嵌入维度 $d \in \{32, 64, 128\}$、层数 $L \in \{2, 3, 4\}$ 和一致性权重 $\lambda_{\triangle} \in [0.1, 10]$）通过在验证子集上进行网格搜索来调优。完整的训练配置和代码见补充材料，并将在论文录用后公开发布。

示例 6.2. 将示例 *4.2* 扩展到具有 $n = 100$ 和$\Omega = \{(i, i+1) \mid i = 1, \ldots, 99\}$ 的链图，令 $a_{i,i+1} = 2$。高效模型对$B_e = 32$ 条边进行子采样，从而生成大小为 ≈ 150 的子图。训练后的得分在留出对上的均方根误差为 $< 0.01 \pm 0.002$，而 *LLS* 基线的均方根误差为0.005。对于 $n = 1000$，稠密训练超出内存限制（> 4*GB*），而我们的版本在标准图形处理器上仅需 < 5 秒即可完成。

6.3.4. 大规模偏好比较矩阵的数值结果

实验在具有真实分数 $x \sim \mathcal{N}(0,1)$ 的合成埃尔多什–雷尼有向图上进行，

$$a_{ij} = \exp(x_i - x_j + \epsilon), \qquad \epsilon \sim \mathcal{N}(0, 0.1),$$

并在 NVIDIA A100 图形处理器上运行。表3比较了运行时间、内存、均方根误差和 Kendall's τ。

表3:

n	p	边	原始运行时间（秒）	原始内存	. (GB)	原始均方根误差	τ	高效运行时间（秒）	高效均方根误差	τ
1000	0.001	499	0.85		0.4	0.312	0.892	0.12	0.315	0.891
1000	0.005	2475	1.24		0.4	0.184	0.951	0.18	0.186	0.950
1000	0.01	4950	1.8		0.4	0.162	0.968	0.22	0.163	0.967
10000	0.001	49950	45.3		4.2	0.194	0.955	1.85	0.195	0.954
10000	0.005	249750	62.1		4.2	0.165	0.972	2.64	0.166	0.971
10000	0.01	499500	78.4		4.2	0.154	0.984	3.12	0.155	0.983
100000	0.001	4999500	OOM		> 32	–	–	18.7	0.182	0.962
100000	0.005	24997500	OOM		> 32	–	–	26.4	0.158	0.978
100000	0.01	49995000	OOM		> 32	–	–	34.2	0.149	0.986

这 5–20× 倍的加速主要源于稀疏内存布局和楔形采样效率，而均方根误差在不同规模下保持稳定（通过增加训练轮次减轻了轻微的随机波动）。这表明该方法能够处理现实世界中的稀疏性，同时不会造成精度下降。

参考文献

[1] R. A. Bradley 和 M. E. Terry。《不完全区组设计的秩分析：I. 成对比较方法》。*Biometrika*，39(3/4)：324–345，1952。DOI：10.1093/biomet/39.3-4.324。[2] R. Duncan Luce。《个体选择行为：理论分析》。John Wiley & Sons，1959。

[3] T. L.Saaty。《层次分析法》。McGraw-Hill International Book Co.，纽约，1980 年。

[4] W. W. Koczkodaj。成对比较一致性的新定义。*Mathematical and ComputerModelling*，18(7):79–84，1993。https://doi.org/10.1016/0895-7177(93)90059-8。

[5] G. Crawford 和 C. Williams。《关于主观判断矩阵分析的札记》。数学心理学杂志，29(4):387–405，1985。https://doi.org/10.1016/0022-2496(85)90002-1。

[6]S. Bozóki、J. Fülöp 和 L. Rónyai。《关于不完整成对比较矩阵的最优补全》。*Mathematical and ComputerModelling*，52(1–2):318–333，2010。https://doi.org/10.1016/j.mcm. 2010.02.047。[7] S. Negahban、S. Oh 和 D. Shah。《Rank Centrality：基于成对比较的排序》。运筹学，65(1):266–287，2017。https: //doi.org/10.1287/opre.2016.1534。[8] F. Fogel、A. d'Aspremont 和 M. Vojnovic。《使用序列排序进行谱排序》。机器学习研究杂志，17(88):1–45，2016。https://www.jmlr.org/papers/v17/16-035.html。[9] Y. Chen、T. Liu 和 S. Oh。《*GNN Rank*：学习基于成对比较的排序》。https://doi.org/10.48550/arXiv.2202.00211，2022年。[10] W. Hamilton、Z. Ying 和 J. Leskovec。《大型图上的归纳式表征学习》。载于第*31*届神经信息处理系统国际会议论文集，2017年12月4日。https://dl.acm.org/doi/proceedings/10.5555/3294771[11] L. Banarescu、M. Abadi、A. Agarwal、P. Barham、E. Brevdo、Z. Chen 等。《TensorFlow：异构系统上的大规模机器学习》。https://arxiv.org/abs/1603.04467。[12] K. C. Agóston 和 L. Csató。《不完整成对比较矩阵的不一致性阈值》。*Omega*，108:102576，2022年。[13] M. Brunelli。《决策分析师为何不应只满足于$n-1$成对比较？来自文献的回响》。载于最佳-最差法进展，运筹学讲义，第33–40页。施普林格，2023年。https://doi.org/10.1007/978-3-031-24816-0_3。[14] W. Cheng、E. Hüllermeier、W. Waegeman 和 V. Welker。《基于阈值概率模型和部分弃权的标签排序》。载于神经信息处理系统进展，第25卷，2012年。https://proceedings.neurips.cc/paper_files/paper/2012/file/fe2d010308a6b3799a3d9c728ee74244-Paper.pdf。