通过矩阵分解的几何解释与分析实现公平推荐

Hao Wanghaow85@live.comRatidar.com 中国北京

摘要

基于矩阵分解的推荐系统本质上是一种保角降维技术。由于物品频率服从幂律分布，原始维度中的大多数用户特征向量和物品特征向量都位于同一个超平面上。然而，从解析角度在原始维度中重建嵌入非常困难，因此我们将原有的保角降维问题重新表述为保距降维问题。我们发现，推荐系统输入数据在原始高维空间中的几何形状分布于具有有趣性质的同心圆上，并设计了一种名为 ParaMat 的基于抛物面的矩阵分解方法来解决推荐问题。在实验部分，我们将该算法与其他 8 种算法进行比较，并证明与 ZeroMat 和 DotMat Hybrid 等当今推荐系统相比，我们的新方法是最公平的算法。

关键词：矩阵分解、ParaMat、线性分解、几何解释、推荐系统 m

1. 引言

如今，每家大型互联网公司都在涉足推荐系统。推荐系统能够大幅提升流量并增加销售收入（在亚马逊，所有销售额中有 30%–40% 来自推荐系统），同时节省大量营销预算。尽管在 2010年代上半期，推荐系统遭遇了重大挫折，当时大多数公司一致认为，推荐系统无法作为独立产品运营；但在悲观论调迅速蔓延后的短短一段时间内，TikTok 和快手等公司便崛起为全球市场上的主要参与者，成为独立的推荐系统产品。

研究人员在几十年前发明了推荐系统，如今，涓涓细流般发展的人工智能技术已汇成洪流。推荐系统的基本理念是运用大数据分析机制分析用户和物品的历史信息，根据计算得出的偏好分数，向用户推荐新物品。例如，如果我们知道 Alice 喜欢阅读苏东坡 、《三国演义》 以及许多其他中国古典文学作品，就可以为她推荐她尚未读过的其他中国古典文学作品。与搜索引擎相比，这种方式更能帮助 Alice 找到新书，因为使用搜索引擎时，Alice 必须自己寻找喜欢的书。我们刚才举例说明的这一过程，是一种名为基于内容的推荐的推荐系统技术的初级形式。它只是推荐系统众多研究领域中的一个子领域。

自2016年以来，RecSys 等知名研究会议见证了深度神经网络模型的兴起。随着越来越多的公司和学校涉足这一领域，推荐系统的技术模型变得越来越复杂，也越来越个性化。由于深度神经网络可以表示任意函数，研究人员得以从比浅层模型时代大得多的候选模型池中选择最合适的模型。然而，由于参数调节的选择众多，关于深度神经网络的公开知识也变得越来越个性化——通常只有少数研究人员了解他们所构建的神经模型背后的原因和原理。

尽管浅层模型时代的早期模型似乎已经过时，但它们仍在许多公司中得到广泛应用。它们经久不衰的一个主要原因是简单且易于解释。基于同样的原因，我们选择矩阵分解作为本文的主要研究对象。我们对经典矩阵分解算法进行几何解释和分析，并基于观察和分析，提出一种名为 ParaMat 的新算法。与其他模型相比，它更易于解释，同时准确度也很高。

2. 相关工作

无论该领域的投资如何起伏，推荐系统仍是一个热门研究领域。作为最成功的推荐范式之一，矩阵分解在 2007 [4] 中以概率框架为基础。该概率框架称为概率矩阵分解。在这一框架的改进基础上，提出了 ZeroMat [5] 和 DotMat [6]，以在没有输入数据的情况下解决冷启动和稀疏性问题。这些算法即使没有历史信息，也能相当准确地预测用户偏好。这两种算法带来的社会影响令人惊叹——人类文化只需经历短短几年演化，便会陷入可预测的状态。

矩阵分解也可用于解决上下文感知推荐（上下文感知推荐系统，CARS）问题 [7] [8]。2021年，推出了一种名为 MatMat [9] 的新型上下文感知推荐系统解决方案。MatMat 不采用标量拟合，而是通过矩阵拟合进行矩阵分解，将上下文信息纳入系统。同年，还提出了 MatMat 的一个实际应用 MovieMat[10]，用于解决电影推荐中的上下文感知推荐系统问题。该算法至多纳入 6 个上下文信息字段，效果优于经典模型。

近年来，公平性成为热门研究课题。谷歌于2017年提出了一种名为 Focused Learning 的算法 [11]，该算法通过公平性指标对矩阵分解损失函数施加惩罚。越来越多的研究人员投入时间和精力研究公平的排序学习方法 [12] [13] [14]，并在 SIGIR 和 KDD 等顶级会议上发表了大量论文。然而，由于矩阵分解简单易用，它仍然是验证公平性理念的良好基准。Zipf 矩阵分解 [15]于2021年提出，同时还引入了一种名为马太效应程度的公平性指标，作为附带成果。MatRec [16]和 KL-Mat [17]也是基于矩阵分解框架的公平推荐算法的例子。2022年，Wang [18] 利用极值理论提出了一组公平性度量。

3. 矩阵分解的几何分析

矩阵分解损失函数最常见的定义之一如下：

$$L = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(R_{i,j} - U_i^T \cdot V_j\right)^2 \tag{1}$$

在该公式中，$_{i,j}$ 表示第 i 个用户对第 j 个物品给出的评分值。$_i$ 是用户特征向量，$_j$ 是物品特征向量。在实际应用中，如果$_i$ 和$_j$ 未经过归一化，我们通常会遇到梯度爆炸问题，因此矩阵分解中更实用的损失函数实际上定义如下：

$$L = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{R_{i,j}}{R_{max}} - \frac{U_i^T \cdot V_j}{||U_i|| \times ||V_j||}\right)^2 \tag{2}$$

由于$U_{iT}\cdot V_i$ $||U_i||\times||V_j||$ 实际上是向量 U_i 与 V_j 之间夹角的余弦，因此我们实际上是在寻找高维空间中的向量，使每对向量之间夹角的余弦值由$R_{i,j}$ R_{max} 定义。

与大多数真实世界的数据集一样，$R_{i,j}$ R_{max} 遵循幂律分布。为简化框架，我们假设这些比率遵循齐夫分布。也就是说，$R_{i,j}$ R_{max} 出现的频率与其自身数值成正比。这带来了输入数据集一个非常有趣的几何性质：$R_{i,j}$ R_{max} 的大多数取值为 1。这意味着，要使公式 (2) 中的 L 最小，大多数 i 和 j 都应共线。

一个自然的问题随之而来：i 和 j 的向量空间是什么样的？我们将共线的 i 和 j 对数定义为 M，那么$M\times R,\ Rmax$ 的向量对的余弦值为$R_{i,j}$ R_{max}。我们发现，若不借助计算机，仅通过解析方法来直观呈现这样的空间，极其困难。

为了考察大多数 i 和 j 的共线性对性质的影响，我们提出如下算法：我们在下面定义新算法（称为线性分解）的损失函数：

$$
\begin{aligned}
L &= \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{R_{i,j}}{R_{max}} - \frac{U_i^T\cdot V_j}{||U_i||\times||V_j||}\right)^2 \\
&\text{Subject to :} \\
&\begin{cases}\sum_{k=1}^{K}\alpha_k\cdot U_{1,k} = 0\\ \cdots \\ \sum_{k=1}^{K}\alpha_k\cdot U_{n,k} = 0\end{cases} \quad and \quad \begin{cases}\sum_{k=1}^{K}\alpha_k\cdot V_{1,k} = 0\\ \cdots \\ \sum_{k=1}^{K}\alpha_k\cdot V_{m,k} = 0\end{cases}
\end{aligned}
\tag{3}
$$

我们在 MovieLens 一百万数据集 [19] (图 1) 上测试了该算法，发现我们提出的算法在技术准确率方面能够取得与其他算法相竞争的结果。根据这一观察，我们可以得出结论：用户和物品特征向量空间的共线性在矩阵分解算法的技术准确率中发挥着至关重要的作用。我们还发现，线性分解是 9 种算法中最公平的推荐系统。我们将在实验部分提供本实验中各算法的参考文献信息。

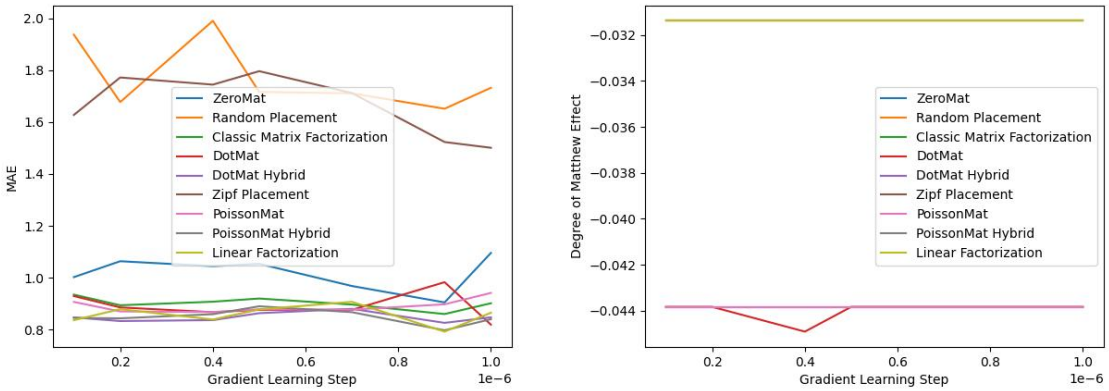


图 1 MovieLens 一百万数据集上的线性分解对比实验（平均绝对误差和马太效应程度）

然而，由于难以设计出适用于所有几何角度保持向量的向量空间泛函，我们采用了另一种途径来解决这个问题。我们不再将 $R_{i,j}$ R_{max} 视为向量之间的余弦角，而是将其视为向量到原点的~~距离~~。这样一来，输入的用户物品评分值（遵循齐夫定律）便成为同心圆上等距分布的样本点。

这一几何性质背后的原因在于，同心圆上等距分布的点的数量与半径长度成正比。如果我们将半径定义为$R_{i,j}$ R_{max}，那么我们新设计的几何中的点便符合齐夫定律。 ——

为了提出名为 ParaMat 的新算法，我们通过以下公式将二维几何扩展到三维空间：

$$z = \frac{R_{i,j}}{R_{max}} \sqrt{x_k^2 + y_k^2} \tag{4}$$

我们将 x 和 y 定义为用户特征向量与物品特征向量的乘积：

$$\begin{aligned} x_k &= U_i^T \cdot V_j \\ y_k &= W_i^T \cdot P_j \end{aligned} \tag{5}$$

ParaMat 的损失函数定义如下：

$$L = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{R_{i,j}}{R_{max}} - d\left(\left(x_k , y_k , \frac{R_{i,j}}{R_{max}} \sqrt{x_k^2 + y_k^2} \right) , (0 , 0 , 0) \right) \right)^2 \tag{6}$$

将 x 和 y 的值替换为用户特征向量与物品特征向量的点积，并将 $R_{i,j}$ R_{max} 替换为经典矩阵分解生成的用户特征向量与物品特征向量的点积。我们得到以下公式：

$$L = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{R_{i,j}}{R_{max}} - \sqrt{\left(U_i^T \cdot V_j\right)^2 + \left(W_i^T \cdot P_j\right)^2 + \left(\frac{R_{i,j}}{R_{max}}\right)^2\left(\left(U_i^T \cdot V_j\right)^2 + \left(W_i^T \cdot P_j\right)^2\right)}\right)^2 \tag{7}$$

我们选择使用随机梯度下降（SGD）算法优化公式（7），然后使用以下公式计算未知的用户—物品评分值：

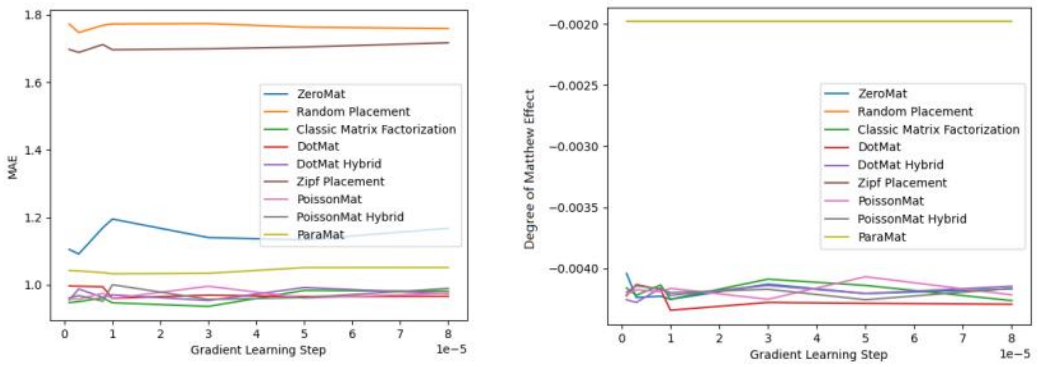
$$R_{i,j} = R_{max}\sqrt{\left(U_i^T \cdot V_j\right)^2 + \left(W_i^T \cdot P_j\right)^2 + \left(\frac{R_{i,j}}{R_{max}}\right)^2\left(\left(U_i^T \cdot V_j\right)^2 + \left(W_i^T \cdot P_j\right)^2\right)} \tag{8}$$

在实验部分，我们展示了 ParaMat 在公平性指标评估下是最佳算法。经过几何分析与推导，我们为推荐系统中的公平性问题找到了一种有效的解决方案。

4. 实验

为了评估 ParaMat 的性能，我们在 MovieLens 一百万数据集 [19] 和 LDOS-CoMoDa [20] 数据集上，将该算法与另外 8 种算法进行了比较。在这些算法中，随机放置表示以均匀随机的方式进行推荐；Zipf 放置表示按照幂律分布进行推荐；ZeroMat [5]和 DotMat[6]在相关工作部分中介绍；PoissonMat [21] 是一种将于 2022年国际会议上发表的算法。

图 2 展示了在 MovieLens 一百万数据集上对 9 种算法进行比较的实验结果。ParaMat 在技术准确率指标平均绝对误差方面排名第 5^{th}，但在公平性指标马太效应程度方面排名第 1^{st}。观察结果表明，按照公平性指标评估，ParaMat 是一种相当有效的推荐系统。由于误差曲线相当杂乱，我们还使用 Takens 嵌入在二维空间中对曲线进行可视化（为便于观察，以三维网格显示）。



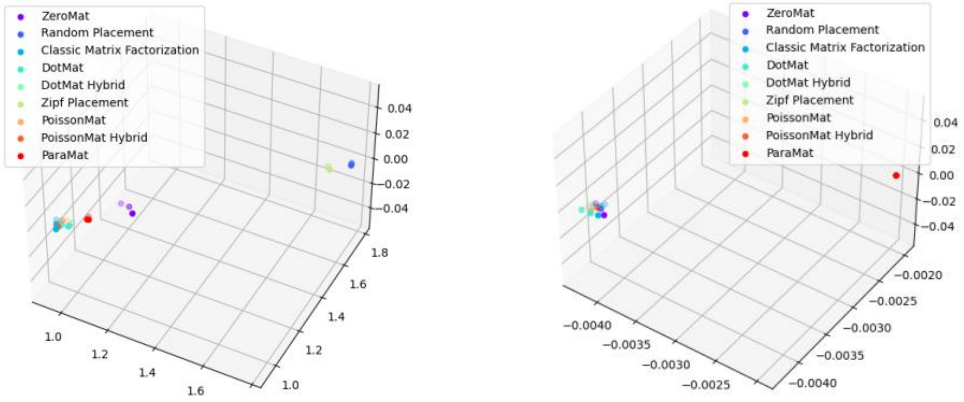


图 2 ParaMat 在 MovieLens 一百万数据集上的比较（平均绝对误差和马太效应程度）。第一行显示了不同算法的平均绝对误差和马太效应程度曲线。底部一行分别使用 Takens 嵌入对这些曲线进行了可视化。

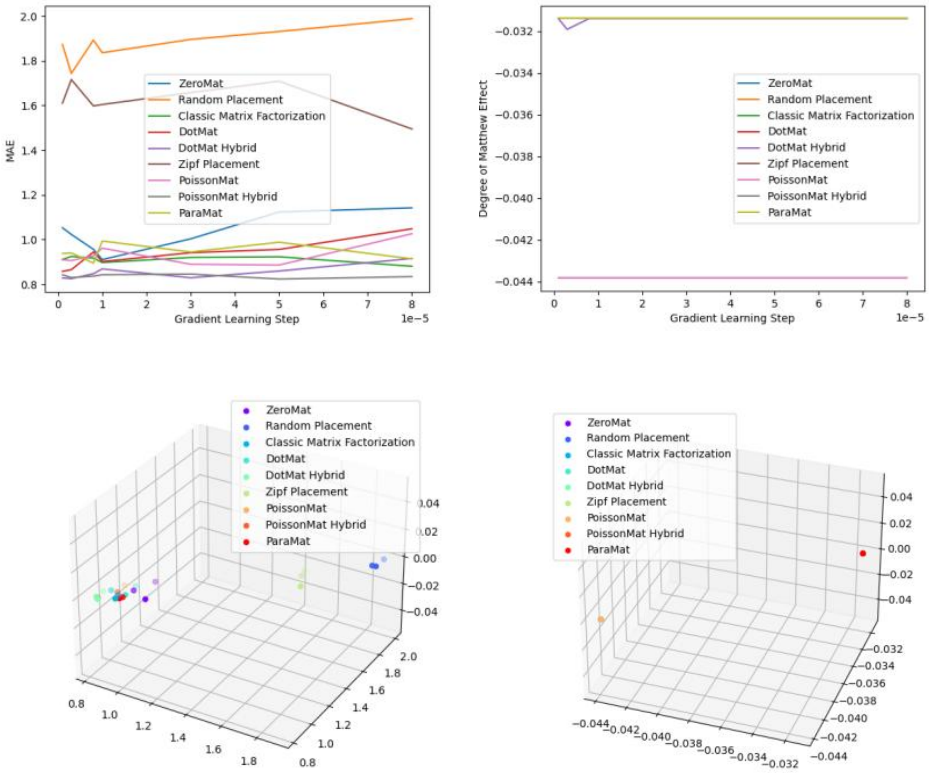


图 3 ParaMat 在 LDOS-CoMoDa数据集上的比较（平均绝对误差和马太效应程度）。第一行显示了不同算法的平均绝对误差和马太效应程度曲线。底部一行分别使用 Takens 嵌入对这些曲线进行了可视化。

图 3 展示了在 LDOS-CoMoDa数据集上对 9 种算法进行比较的实验结果。按照技术准确率指标平均绝对误差（MAE），ParaMat 排名第 4^{th}；但按照公平性指标马太效应程度，ParaMat 再次排名第 1^{st}。观察结果可知，ParaMat 是最公平的推荐系统。通过 Takens 嵌入，可以在二维空间中更清晰地呈现这一结果。

5. 结论

本文提出了两种几何驱动的公平推荐系统算法，分别命名为线性分解和 ParaMat。线性分解基于这样的观察：用户特征向量和物品特征向量大多共线，因此假设矩阵分解中的用户特征向量与物品特征向量位于同一超平面上。ParaMat 同样基于几何观察，即用户物品评分值大多等距地分布在同心圆上。

线性分解和 ParaMat 在平均绝对误差等技术准确性指标上的表现都不是最佳，但在实验部分评估的9种算法中，它们是最公平的算法。通过分析输入数据结构的几何空间，我们提出了两种有效的公平推荐算法。

在未来的工作中，我们希望探索其他经典算法（例如排序学习和因子分解机）的输入数据结构几何空间。我们也对深度学习模型的几何解释很感兴趣。

参考文献

[1] H. Guo, R. Tang等，“DeepFM：基于因子分解机的点击率预测神经网络”，IJCAI，2017年[2] G. Zhou, X. Zhu, C. Song等，“用于点击率预测的深度兴趣网络”，KDD，2018年[3] R. Mehta; K. Rana，“推荐系统中的矩阵分解技术综述”，CSCITA，2017年[4]R.Salakhutdinov, A.Mnih，“概率矩阵分解”，第20届神经信息处理系统国际会议论文集，2007年[5] H. Wang，“ZeroMat：利用无输入数据解决推荐系统冷启动问题”，IEEE第4届信息系统与计算机辅助教育国际会议（ICISCAE），2021年[6] H. Wang，“DotMat：解决推荐系统冷启动问题并缓解稀疏性问题”，IEEE第5届电子技术国际会议（ICET 2022）[7] E. Coviello, K. Ellis等，“可扩展的在线上下文音乐推荐模型”，CARS研讨会，2021年[8] Y. Chen, J. Li，“为下一篮子推荐建模动态属性”，CARS研讨会，2021年[9] H.Wang，“MatMat：通过矩阵拟合进行矩阵分解”，IEEE第4届信息系统与计算机辅助教育国际会议（ICISCAE），2021年[10]H.Wang，“MovieMat：通过矩阵拟合进行矩阵分解的上下文感知电影推荐”，第7届计算机与通信国际会议（ICCC），2021年[11]A. Beutel, Ed. Chi等，“超越全局最优：通过聚焦学习改进推荐”，WWW，2017年[12]H.Yadav, Z.Du, T.Joachims. “基于隐式反馈的公平学习排序。”，SIGIR，2020年[13]M.Morik, A.Singh, J.Hong, T.Joachims. “动态学习排序中的公平性与偏差控制”，SIGIR，2020年[14]M.Zehlike, C.Castillo. “减少排序中的差异化曝光：一种学习排序方法。”，SIGIR，2020年[15]H. Wang，“Zipf 矩阵分解：缓解马太效应的矩阵分解”，ICAIBD，2021年[16]H. Wang, B. Ruan，“MatRec：针对高度偏斜数据集的矩阵分解”，ICBDT，2020年[17]H. Wang，“KL-Mat：通过信息几何实现公平推荐系统”，icWCSN，2022年[18]H. Wang，“推荐系统的公平性度量”，icWCSN，2022年[19]T. Bertin-Mahieux, B. Whitman, P. Lamere，《百万歌曲数据集》，ISMIR，2011年[20]ODIĆ, Ante, TKALČIČ, Marko, TASIČ, Jurij F., KOŠIR, Andrej：《预测和检测电影推荐系统中的相关上下文信息》，Interacting with Computers，第25卷，第1期，2013年[21]H. Wang，“PoissonMat：使用泊松分布重构矩阵分解，并在无输入数据的情况下解决冷启动问题”，即将发表于 MLISE，2022年