**Social Recommendation With Local Low Rank Matrix Approximation**

# Abstract

在推荐文献中，社交关系已成功融入传统的推荐方法中，尤其是最受欢迎的状态矩阵分解（Matrix Factorization，MF）。**MF基于用户项目评分矩阵为低秩的假设**，因此可以分解decomposed为两个代表用户和项目潜在特征的较小矩阵。最近，不是假设评级矩阵是低秩，而是基于**一种新的假设提出了局部低秩矩阵近似**（LLORMA），该假设是评级矩阵由一组低秩子矩阵组成，称为局部低秩。不是将原始矩阵分解，而是将这些低排序子矩阵独立分解以近似原始评分矩阵。实验结果表明，与标准MF相比，LLORMA显着提高了推荐性能。**不幸的是，LLORMA不利用任何社交信息。**在本文中，我们提出了一种称为社会局部加权矩阵分解（SLWMF）的新模型，这是第一个在局部低阶框架中对社会推荐进行建模的工作。**SLWMF基于这样的直觉（直观的）的想法based on the intuitive idea：如果我们可以智能地将社交关系整合到局部的低级框架中，那么推荐性能肯定会得到改善**。通过假设存在一组有影响力的用户（称为连接器），其作为社交图的基础，谁可以在至少三度内影响人的行为，我们构建以一组有影响力的用户为中心的子矩阵，而不是LLORMA中使用的随机点。我们认为**SLWMF优于LLORMA的一个主要优点是前者可以获得比后者更有意义的子矩阵**。此外，我们还设计了比LLORMA更有效的优化目标函数，并对两个真实数据集进行了广泛的实验。实验结果表明，SLWMF可以进一步提高推荐性能与几个最先进的基线相比。

# INTRODUCTION

推荐系统（RS）已经成为大数据时代处理信息过载问题的强大工具。一般来说，RS旨在帮助用户根据以前的行为获得他们感兴趣的内容。协作过滤（CF）是RS的一种最先进的方法，它尝试根据类似的用户或项目预测用户对看不见项目的评分（或偏好）。例如，在Netflix，用户将在观看后给予评级（从1到5）到电影。那么Netflix会根据他或她的评级历史来预测用户对未看过的电影的评分。CF的基本假设是人们会喜欢与他们过去喜欢的东西类似的东西，而类似的用户也有类似的喜好。CF方法可以大致分为基于记忆和基于模型的方法。基于记忆的方法主要利用用户项目评分矩阵中用户或项目的邻域信息，而基于模型的方法试图利用系统技术来发现和预测用户的偏好。尽管基于记忆的方法更简单，但基于模型的方法具有更强的预测能力。在各种基于模型的方法中，矩阵分解（MF）由于其在大型数据集中的良好性能和可扩展性而成为近年来最流行的方法[17,16,6,7]。MF的假设是用户对项目的偏好是由一小组潜在因素控制的。因此，大用户项目评分矩阵可以分解成两个较小的矩阵，代表用户特定的和特定项目的潜在因素。换句话说，原始矩阵拥有低阶属性。

尽管表现良好，MF仍然受到冷启动问题和评分矩阵的稀疏性影响。为了解决这些问题，基于假设我们的行为和偏好受到我们的朋友在社交网络中的影响，MF框架中引入了社交信息。在[14,13,5,15,11]中已经提出了几种最先进的方法，其中大多数方法明确地关注直接朋友在社交图谱中的影响。对于稀疏性问题，一个潜在的方向是将原始矩阵分解成一组更小但密度更高的子矩阵，然后将这些子矩阵中的额定值矩阵近似。Lee等人提出了局部LOw秩矩阵近似（LLORMA），与标准MF相比，其显着提高了推荐性能。在LLORMA中，Lee等人不是假设原来的评级矩阵是低级别，称为全球低级别，假定它是由许多子矩阵组成的，这些子矩阵拥有低秩属性，称为局部低秩。LLORMA是基于这样一种动机，即实际上往往只有少数人对少数项目感兴趣，从而形成了子矩阵的结构。已经设计了一个系统的方法来从等级矩阵构造子矩阵。因为子矩阵具有低秩属性，所以标准MF框架可以很容易地独立应用于它们。最后，它们被合并以近似原始评分矩阵。

LLORMA的成功证明了局部低秩的有效性。事实上，在其他应用中已经开发了局部低秩假设的想法，并且也实现了最先进的表现。例如，在彩色化[22]中，不是假设整个图像是低秩的，而是从大图像的局部区域提取的一组相似斑块应该是低秩的。然后，在多标签学习中，提出了一种局部嵌入方法，其中低秩假设仅施加于相似标签的子集[1]。局部低秩框架的关键部分是如何构建一个拥有原始矩阵的低秩属性的子矩阵。因此，如果我们能够将社交关系智能地融入到LLORMA框架中，我们就可以通过一个直观的想法激发我们进一步提高推荐绩效。在本文中，我们提出了一种称为社会局部加权矩阵因子分解（SLWMF）的新模型，它是第一个将社交关系并入LLORMA框架的系统方法。此外，我们认为，在社会关系的帮助下，LLORMA框架的主要弱点可以得到显着缓解。此外，我们观察到LLORMA的主要弱点在于子矩阵的构造。具体来说，它首先从等级矩阵中随机选择若干个用户/项目对，称为定位点，然后根据用户和相邻项与定位点之间的项目相似度为定位点选择邻居。然而，我们声称随机选择锚点的方法是临时的，因此不能保证产生有意义的子矩阵。

在SLWMF中，子矩阵以新颖而有意义的方式构建。根据众所周知的三度影响力（TDI），我们假设存在一组有影响力的用户，称为连接者，他们可以将影响力发挥到社交图中至少与他们三度之内的人。识别连接器后，可轻松选择连接器可传播影响的人群。然后，可以在连接器上构建一个子矩阵，以及圆圈中的人员以及他们与之交互的项目。为了模拟连接器在不同程度上对人们施加影响的各种优势，我们设计了一种启发式方案来衡量影响的强度。此外，为了同时捕捉直接朋友在社交图谱中的影响，我们将它们建模为社会正则化术语，这是将社会推荐纳入标准MF的常见做法。然后，每个子矩阵通过具有社会正则化的加权矩阵分解来近似，称为局部社会正则化。最后，原始矩阵中的缺失预测可以通过包含用户和项目的所有子矩阵的加权平均来预测。

# 3. PROPOSED FRAMEWORK

## 3.1 Motivation

推荐系统的主要目的是根据用户的喜好向用户推荐商品。但是，用户的偏好可能会受到其他人的影响，特别是具有大量连接的有影响力的用户。随着社交网络的激增，世界上的每个人都彼此接近。这种小世界现象3，也与着名的六度分离（SDS）有关，意味着人们容易受到影响并影响其他人。更重要的是，众所周知，社交网络上的一些有影响力的用户会对很多人产生很大的影响。有影响力的用户拥有大量的直接连接（例如，Facebook上的大量朋友或Twitter上的追随者）。根据TDI理论，这种现象对于推荐非常重要。

基于上述动机，我们提出的方法通过考虑用户所属的圈子以及来自他或她的直接朋友的圈子的所有影响来预测用户对项目的偏好。为了纳入这两种社会影响，从LLORMA延伸出来，我们提出了一种称为社会局部加权矩阵分解（SLWMF）的新模型，其中我们假设原始评分矩阵可以用一组子矩阵表示，每个子矩阵都是基于在隐藏的圈子上，即用户和他们在隐藏圈子内相互作用的项目。此外，如[9]中所述，我们假设所有的子矩阵都具有低秩属性。因此，我们可以将矩阵分解技术应用于所有子矩阵，以获得与标准MF框架中相同的用户和项目的潜在特征。然后，可以通过组合来自包含用户和项目的子矩阵的所有预测来预测用户对项目的评级。

算法过程：

1.识别社交图G中最有影响力的用户，称为connectors。

2.构建以q个connectors为中心的子矩阵。

3.设计一个优化框架，将所有子矩阵分解为 .

4.通过组合来自所有子矩阵的预测来预测原始矩阵R中缺失的评级。

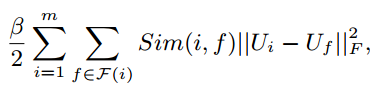
**3.3 Connectors Identification**

连接器是圈子中影响力的来源。因此，我们需要确定社交网络中最有影响力的用户。事实上，这与社交网络中众所周知的影响最大化问题相对应，该问题试图在社交图中找到一小组有影响力的用户，以使他们能够最大化影响力的传播。影响最大化问题已经在文献中广泛存在。在本文中，我们采用了流行作品中提到的启发式方法[3]，其中Even-Dar和Shapira表明影响最大化的最佳种子只是最高度的节点，即那些在社交图中拥有最多朋友的人。具体而言，q连接器的集合Nq由以下公式获得

其中f（u）表示用户u在社交网络中的朋友的数量。事实上，我们采用这种启发式方法来考虑更多的问题：1）它与直觉相符合，即拥有更多朋友的人倾向于具有更具吸引力的特征，因此人们更可能调整他们的行为或偏好; 2）就小世界现象而言，格拉德威尔认为SDS依赖于少数拥有大量朋友网络的非凡人物，他们称之为连接者; 3）这是一种启发式但有效的方法，符合奥卡姆剃刀的规则 实验结果证明了它的有效性。

## 3.5 Local Social Regularization

基于社会同质性，我们认为，在每个子矩阵，越close的直接朋友，隐特征就会越相似。因此，与[15]中使用的类似，我们按照以下方式在用户的隐特征上添加社会正规化：



其中β> 0，F（i）是ui的好友集合，Sim（i，f）是测量ui和uf之间相似度的相似度函数，也就是说ui和uj越接近，它们的潜在因子就越接近 应该Ma等人。将这些正则化术语纳入全球MF框架[15]，而在我们的工作中，我们将它们整合到子矩阵分解过程中。为了计算两个用户之间的相似度，我们利用流行的人际相关系数（PCC）:

