Mixed factorization for collaborative recommendation with  
heterogeneous explicit feedbacks

# ABSTRACT

协同推荐（CR）是在各种在线和离线应用中提供高质量个性化服务的基本支持技术。具有异质显式反馈（CR-HEF）的协同推荐（诸如五星级评分和类似/不喜欢的二进制评级）是一个新的而且重要的问题，因为它为学习用户的偏好提供了丰富而准确的来源。然而，大多数以前关于协作推荐的研究只关注于利用同质的显式反馈，如等级分数或同类隐式反馈，如点击或购买。在本文中，我们研究了CR-HEF问题，设计了一个新颖的，通用的基于混合因子分解的迁移学习框架，以充分利用这两种不同类型的显式反馈。两个CR-HEF任务与实际数据集的实验结果表明，我们的TMF能够比最先进的方法有更好的表现。

# 1. Introduction

协同推荐[1,2]作为一种在各种系统和应用中提供个性化服务的支持技术，如学术资源推荐[3,4]，娱乐视频推荐[5,6]，电信/移动服务推荐[7] 8]和人/社区推荐[9,10]等协同推荐的主要思想是通过集体的方式利用用户的反馈来学习用户的隐藏偏好，而不是分别研究每个用户[11,12]。然而，当用户的反馈很少时，我们可能不会很好地学习用户的偏好，特别是大多数推荐算法[15,16]所依赖的五星级评分。这种数据稀缺的挑战通常会在建立推荐模型时引起过度拟合问题。

异质显式反馈（HEF），如从最好到最差的等级分数以及喜欢和不喜欢的二进制等级，为学习用户偏好和构建用户在线个人资料提供了丰富而准确的来源，这为我们提供了解决数据稀缺性挑战的机会的成绩[17]。然而，协同推荐（CR）中的大多数数学模型是为了从一致的显式反馈（如五星级评分[15,16]）或同类隐式反馈（如点击或购买）[18,19]中学习用户偏好而设计的。在本文中，我们研究这个新的和重要的问题，即包含两种不同类型的显式反馈（即五星级评分和类似/不喜欢的二进制评级）的异质显式反馈协同推荐（CR-HEF）。然后，我们着重于设计一种新颖而通用的算法，以便以原则的方式充分利用这种异构的显式反馈。

以前关于利用异构显式反馈的研究很少，其中主要的方法是矩阵集体分解[17,20,21]。基于集体分解的方法通常被设计为同时分解两个偏好矩阵，即CR-HEF中的等级评分矩阵和二元评分矩阵，其中共享来自相同用户或相同项目的一些潜在特征以便实现联合偏好建模和共享。然而，这两个联合执行的因子分解任务仍然是松散耦合的，可能无法充分利用二进制评分来评分。

除了集体分解方法之外，还有一种称为**矩阵集成因子分解**的方法，用于CR-HEF的相关问题，即具有等级分数和隐式反馈的协同推荐[15]。基于综合因子分解的方法将隐式反馈作为实例（或反馈实例），并将它们无缝地结合到等级分数的分解任务中，作为预测规则中的一个附加术语。**然而，以这种综合的方式利用隐含的反馈到等级分数可能不能很好地捕捉隐式反馈依赖效应。**

在本文中，我们的目标是克服上述两个最先进的因子分解模型的局限性，即**集体因子分解和一体化因子分解**，并使之适应我们所研究的CR-HEF问题。具体而言，我们首先从迁移学习视角[22]中选取CR-HEF问题，其中以成绩评分为目标数据，以二进制评分作为辅助数据。然后提出了一种基于混合因子分解的新型混合分解迁移学习框架，即混合分解迁移（TMF），即集合分解和一体化因子分解作为不同的组装分量。我们的TMF的新颖之处在于它将集体因子分解和综合因子分解统一在一个单一的迁移学习框架中，使基于特征的和基于实例的偏好学习和原则迁移成为可能。此外，我们还可以从我们的TMF推导出一些新的算法变体，以利用不同的部分或辅助二进制评级的组合。

TMF有望将更多的知识从二进制评分迁移到等级分数，而不是集体因素分解，并且比一体化因子分解更准确地模拟二元评级依赖和非独立效应。两个真实世界的数据集上的实验结果表明，我们的TMF比单独的集体分解或集成因子分解显着地好。

我们按以下方式组织其余的论文。我们首先在第二部分讨论一些密切相关的作品。然后，我们在第3节中详细描述所提出的解决方案，并在第4节中对我们的TMF和最新的方法进行广泛的实证研究。最后，我们在第5节给出一些结论性意见和未来的方向。

# 2. Related work

我们的TMF是从具有分解技术的转移学习角度设计的，旨在学习用户的偏好来自协同推荐中的异构显式反馈（HEF）。 因此，在本节中，我们讨论一些相关的问题TMF在异构反馈转移学习环境下的工作以及协同推荐的分解。

## 2.1. Transfer learning for heterogeneous feedbacks

转移学习[22]旨在为多个领域，任务或数据进行参数学习和知识转移，这在许多应用中都是最先进的解决方案，其中包括文本和图像分类[23,24]，个性化和社会行为学习[17,25]等转移学习算法也被设计为在协同推荐中从不同的反馈中学习用户的偏好，为此我们将讨论他们的问题设置和技术。

异构反馈的问题设置主要包括两类，即（i）具有相同反馈类型的不同推荐任务，以及（ii）具有不同反馈类型的相同推荐任务。第一类包括书籍和电影的评分推荐任务[13,26]，书籍推荐任务，电影和音乐评分[27]，推荐任务不同的亚马逊产品评分[28,29]，等等。这个类别通常被称为跨域协作推荐[30,31]，因为项目或产品领域是不同的。第二类包括推荐使用不同类型反馈的电影，如评分和隐式，二进制和/或不确定的反馈[14,17,32]。本文研究的CR-HEF问题也属于第二类。在协同推荐或其他应用中，异构反馈的转移学习技术主要包括基于模型的，基于特征的和基于实例的转移[22,32]，相应的算法包括自适应，集体和集成风格。在[32]中，我们遵循并扩展了协同推荐中转移学习技术的分类，特别是转移学习中的“如何转移”[22]，并展示了我们的TMF与表1中的一些典型作品之间的关系。从表1可以看出，我们的TMF不同于现有的所有作品。这两个维度中的任何一个，即转移学习方法或转移学习算法风格。具体而言，TMF是基于特征的转移和基于实例的转移的混合转移学习方法，以及集体分解和综合因子分解的混合转移学习算法。

## 2.2. Factorization for collaborative recommendation

分解技术，如二阶矩阵分解，高阶张量分解及其扩展已经很好地研究并成功应用于许多机器学习和数据挖掘问题[33]，其中协作推荐是一个重要的应用[15,16,19,34-37]。