Fast Matrix Factorization for Online Recommendation with Implicit Feedback

# ABSTRACT

本文对隐含反馈矩阵分解（MF）方法的有效性和效率进行了改进。我们强调现有工作的两个关键问题。首先，由于没有观察到反馈的空间很大，大多数现有的工作都要为丢失的数据赋予一个统一的权重，以减少计算的复杂性。然而，这种统一的假设在现实世界中是无效的。第二，大多数方法也是在离线设置中设计的，不能跟上在线数据的动态性质。

我们从隐含的反馈中解决学习MF模型的上述两个问题。我们首先提出基于项目流行度减少丢失的数据，这比均匀权重假设更有效和灵活。然而，这种不均匀的加权在学习模型方面构成了效率挑战。为了解决这个问题，我们专门设计了一种基于元素交替最小二乘法（eALS）技术的新的学习算法，用于有效优化具有可变加权丢失数据的MF模型。我们利用这种效率，然后无缝地设计一个增量更新策略，立即刷新MF模型给出新的反馈。通过对离线和在线协议中的两个公共数据集进行全面的实验，我们可以看出，我们实现的，开源的eALS始终优于最先进的隐式MF方法。

# INTRODUCTION

用户个性化在现代推荐系统中已经普及。它有助于捕获用户的个性化偏好，并且已经显示增加用户对内容提供商的满意度和收入。在各种方法中，矩阵分解（MF）是通过潜在因子向量来表征用户和项目的最流行和最有效的技术[15,32]。关于MF算法的早期工作推荐[14,27]主要集中在明确的反馈，用户的评级直接反映了他们对项目的偏好。这些作品将推荐作为一个评级预测问题，假设大量未观察到的评级（即，丢失的数据）对用户偏好的建模是无关的[12]。这大大减少了建模工作量，并且已经设计了许多复杂的方法，例如SVD ++ [15]和timeSVD [14]。

然而，在许多应用中，显性评级并不总是可用; 更经常地，用户通过隐含的反馈（例如，用户的视频观看和产品购买历史）与项目进行交互。与明确的评级相比，隐含的反馈对于内容提供商来说更容易收集，但由于**负面反馈的自然稀缺性**，更具有挑战性。已经表明，仅建模观察到的正反馈导致用户简档中的偏见表示[4,12]。 例如，Marlin等人[18]发现用户听他们期望喜欢的音乐，并避免他们不喜欢的类型，导致观察数据的严重偏差。

为了解决缺乏负反馈的问题（也称为一级问题[21]），一个流行的解决方案是将**所有丢失的数据**建模为负反馈[12]。然而，由于充分考虑了观察和丢失的数据，这会不利地降低学习效率。更重要的是，低效率使得在线部署隐式MF方法更加困难[29]。在实际推荐系统中，新用户，项目和交互连续流式传输，实时刷新底层模型至关重要，以最好地为用户提供服务。在这项工作中，我们关注MF方法的上述两个具有挑战性的问题 隐含反馈和在线学习。我们注意到，作为Devooght等人最近的一项工作，我们不是首先考虑MF的两个方面。[4]提出了一种用于动态数据学习的有效隐含MF方法。然而，我们认为Devooght的方法[4]以不现实的，次优的方式对missing数据进行建模。然而，这样的假设限制了模型对实际应用的保真度和灵活性。例如，内容提供商通常知道哪些项目已被用户频繁显示，但很少被点击; 这些项目更可能是真正的负面评估，应该被加权高于其他项目。此外，Devooght的方法通过梯度下降来学习参数，需要昂贵的线搜索来确定每个步骤的最佳学习速率。

我们提出了一种新的MF方法，旨在有效地学习隐式反馈，同时满足在线学习的要求。我们开发了一种新的学习算法，可以有效优化隐含的MF模型，而不会对missing数据施加统一的权重限制。特别地，我们基于项目的受欢迎程度来分配missing数据的权重，这可以比以前的受限于均匀性假设的方法[4,12,23,30,31]更有效。我们的eALS算法能够快速识别缺失数据，分析速度比ALS快K倍[12]，其中K表示潜在因子的数量，与最近的动态MF解决方案相同的时间复杂度[4]。这种效率水平使得eALS适合在线学习，为此我们开发了一种增量更新策略，可立即刷新给定新输入数据的模型参数。eALS的另一个关键优点是，它无需学习速率，绕过众所周知的用于调整梯度下降方法（如[4]和随机梯度下降（SGD）[25]）的困难。我们总结我们的主要贡献如下。

1.我们提出了一个完整的缺失数据的项目流行度感知加权方案，有效地定制了MF模型，用于从隐式反馈中学习。

2.我们开发了一种新的算法，有效地学习模型参数，并制定了增量更新策略来支持实时在线学习。

3.我们在两个现实世界的数据集上对离线和在线协议进行了广泛的实验，表明我们的方法始终优于现有技术的隐含MF方法。

# 2. RELATED WORK

由于缺乏负面反馈，处理缺失数据是隐含数据学习的必要条件。为此，已经提出了两种策略，基于样本的学习[21,25]从丢失的数据中抽取负面实例，或全数据学习[12,30]将所有缺失的数据视为负数。这两种方法都有利弊：基于样本的方法通过减少训练中的负面例子而更有效，但风险降低了模型的预测性; 基于整体的方法将完整的数据模拟成可能具有较高的覆盖范围，但效率低下可能是一个问题。为了保持模型的保真度，我们坚持在整个数据基础的学习，发展快速的基于ALS的算法来解决效率低下的问题。

对于现有的基于全数据的方法[4,12,23,30,31]，一个主要的限制是对missing data的统一加权，这有利于算法的效率，但限制了模型的灵活性和可扩展性。考虑不均匀加权的唯一工作来自Pan等[20,21]。然而它们的立方时复杂度w.r.t. K使其不适合运行大规模数据[23]，其中需要考虑大量因素来获得改进的性能[31]。

为了优化MF，已经对各种学习者进行了调查，包括SGD [14,25]，Coordinate Descent（CD）[4,32]和Markov Chain Monto Carlo（MCMC）[26]。由于导出的容易性SGD是最流行的一个，然而，由于大量的训练实例（全用户项目交互矩阵被考虑），它不适用于基于全数据的MF [12]。**ALS可以被看作是CD的实例化，已被广泛用于解决整个MF** [12,20,21,30]。 然而，它的低效率是实际使用的主要障碍[23,31]。为了解决这个问题，[23]描述了ALS的一个近似解决方案。最近，[4]采用随机块坐标下降（RCD）学习者[28]，降低复杂性并将其应用于动态场景。类似地，[31]类似地丰富了基于邻居的隐式反馈矩阵，然后应用未加权的SVD。与以前的作品不同，我们提出了一种基于全数据的MF的非均匀缺失数据的有效元素ALS解决方案，以前从未研究过。实用推荐系统的另一个重要方面在于处理传入数据的动态性质，其及时性是关键考虑因素。由于无法在线重新整理完整的模型，各种作品已经为基于邻居的[13]，基于图的[9]，概率[3]和MF [4,5,17,27]的方法开发了增量学习策略。对于MF，已经研究了不同的学习者，用于在线更新，包括SGD [5,27]，RCD [4]和双重分割[17]。据我们所知，这项工作是首次尝试利用ALS技术进行在线学习。

实用推荐系统的另一个重要方面在于处理传入数据的动态性质，其及时性是关键考虑因素。由于无法在线重新整理完整的模型，各种work已经为基于邻居的[13]，基于图的[9]，概率[3]和MF [4,5,17,27]的方法开发了增量学习策略。对于MF，已经研究了不同的学习者，用于在线更新，包括SGD [5,27]，RCD [4]和双重分割[17]。据我们所知，这项工作是首次尝试利用ALS技术进行在线学习。

# PRELIMINARIES

ALS算法：<http://blog.csdn.net/oucpowerman/article/details/49847979>

<http://www.jianshu.com/p/9a584bba1c68>

由于评分数据中有大量的缺失项，传统的矩阵分解SVD（奇异值分解）不方便处理这个问题，而ALS能够很好的解决这个问题。

我们首先介绍基于**全数据的MF方法来从隐式数据学习**，**突出传统ALS解决方案的低效问题**[12,21]。然后，我们描述eALS，一个元素方面的ALS学习者[26]，可以将线性度的时间复杂度减少到多个因素。虽然学习者在通过各种加权策略优化MF是通用的，但是对于所有缺失的数据来说，引入的形式是昂贵的，因此实际使用是不切实际的。这个缺陷促使我们进一步开发eALS，使其适合于从隐含的反馈中学习（详见第4.2节）。

## MF Method for Implicit Feedback

矩阵因式分解将用户和项目映射到K维度的关联潜在特征空间中，使得交互被建模为该空间中的内部产品。在数学上，R的每个入口rui估计为：



## Generic Element-wise ALS Learner

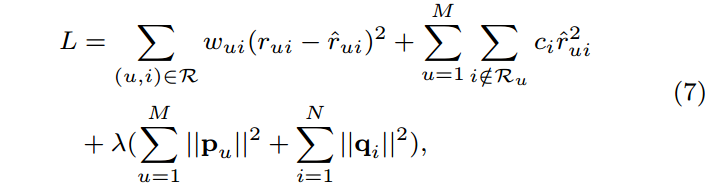
以前的ALS解决方案的瓶颈在于矩阵求逆运算，这是由于设计为整个用户（项目）更新潜在向量。因此，在元素级优化参数是很自然的：优化潜在向量的每个坐标，同时保留其他坐标[26]。为了实现这一点，我们首先得到关于puf的目标函数方程（2）的导数.

# OUR IMPLICIT MF METHOD

我们首先针对缺失的数据提出一种面向项目的加权方案，并遵循受欢迎程度的加权策略，这可以比推荐任务的统一加权更有效。然后，我们开发了一种快速的eALS算法来优化目标函数，与常规ALS [12]和通用元素ALS学习者相比，显着降低了学习复杂性[26]。最后，我们讨论如何调整实时在线学习的学习算法。

## Item-Oriented Weighting on Missing Data

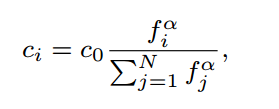
由于项目空间大，用户缺少的条目是**负面和未知反馈的混合**。在指定缺少条目的权重时，希望为负反馈分配更高的权重。然而，区分这两种情况是一个众所周知的困难。另外，由于相互作用矩阵R通常是大而稀疏的，所以存储每个零条目个体化权重将是耗费巨大的。为此，现有的工作[4,12,23,30,31]对缺少的条目应用了一个简单的均匀权重，然而对于实际应用来说，它们不是最佳的和不可扩展的。考虑到内容提供商**访问项目侧的负面信息的便利性**（例如哪些项目已被推送给用户但几乎没有互动）， 我们认为，根据某些项目属性来减少缺失的数据更为现实。为了捕捉这个，我们设计了一个更精细的目标函数如下：



其中ci表示用户“错过”的项目是**真正的负面评估的信心**，这可以作为从实践者编码领域知识的手段。很显然的是，第一项表示所观察到的条目的预测误差，这在显式建模评分[15，27]被广泛采用。第二项占丢失的数据，其作为负实例的作用，并且是用于从隐式反馈[12，25]建议是至关重要的。接下来，我们提出了一个独立于领域的策略，通过利用现代Web 2.0系统的普遍特征来确定ci。

### Popularity-aware Weighting Strategy 流行度感知加权策略

许多Web 2.0系统的现有视觉界面在其建议中展示了热门项目。所有其他因素相同，受欢迎的项目更有可能被用户普遍认识[10]。因此认为**一个受欢迎的项目的“缺失”**对于用户来说更有可能是**真正无关紧要的（而不是未知的）**。为了解决这个问题，我们根据项目的受欢迎程度参数化Ci:



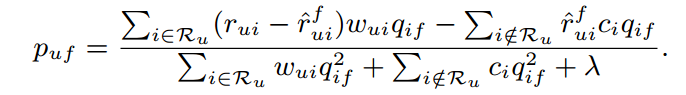
α>1时，受欢迎的物品的权重被促进去加强与不受欢迎的商品的差异;而设置α在（0，1）的较低范围内，可以抑制流行商品的权重，并具有平滑效果。 我们经验发现α= 0.5通常导致良好结果。 请注意，统一加权是一种特殊情况将α设置为0。

Relationship to Negative Sampling

我们提出的流行度加权策略与Rendle基于流行度的**过采样**[24]具有相同的直觉，用于学习BPR，其基本上以较高的概率将受欢迎的项目作为负反馈。然而，[24]经验表明，**过采样**方法表现优于基本均匀采样器。我们怀疑来自SGD learner的原因，这将导致流行项目更多的梯度步骤，由于过度抽样。因此，受欢迎的项目可能在本地过度训练，而不受欢迎的项目受到不足的训练。为了解决这个问题，其他领域已经采用了抽样频繁项目[19]和自适应学习速率（如Adagrad [6]）的技巧。由于本文的重点是基于全数据的隐含MF，我们不再进一步探索SGD的细节。值得指出的是，我们提出的eALS学习者通过对每个模型参数的精确优化来避免这些学习问题。

## Fast eALS Learning Algorithm

我们可以通过**避免加权缺失数据引入的大量重复计算**来加快学习速度。我们详细介绍了puf的推导过程 同样地实现了qif的对应物。首先，我们通过分离观察到的数据部分来重写puf更新规则Eq（5）：



显然，计算瓶颈在于丢失数据部分的求和，这需要遍历整个负空间。 我们首先关注分子：