**基于二部网络表示学习的矩阵分解推荐算法**

袁梦祥*1*，张以文*1*，颜登程+*2*，周珊*3*

(1.安徽大学计算机科学与技术学院，安徽 合肥 230601；

2.安徽大学物质科学与信息技术研究院，安徽 合肥 230601；

3. 深圳易伙科技有限责任公司，广东 深圳 518000)

**摘要：**矩阵分解算法广泛应用于推荐系统。然而，其性能往往受到数据稀疏性和数据高维度的影响，且较少考虑项目的内容信息。针对上述问题,本文提出了一种联合二部网络表示学习的矩阵分解推荐算法(BiNRMF)。首先，利用评分信息和项目的标签信息构建两个二部网络；然后，通过二部网络的表示学习算法得到用户和物品的低维向量表示，用以计算用户之间和物品之间的相似性；最后，改进传统矩阵分解模型，融入低维向量空间中用户的相似关系和物品的相似关系。在 GoodBooks 和 MovieLens 数据集上的实验结果表明与经典的协同过滤和矩阵分解算法相比，该算法在预测精度上有显著提升。

**关键词：**推荐系统；协同过滤；二部网络；网络表示学习；矩阵分解

**中图分类号: 文献标识码：**

**Regularizing Matrix Factorization with Bipartite Network Representation Learning for Recommendation**

ZHANG Yi-Wen1，YUAN Meng-Xiang1，YAN Deng-Cheng2

1.School of Computer Science and Technology, Anhui University, Hefei 230601

2. Institute of Physical Science and Information Technology, Anhui University, Hefei 230601

**Abstract:** Matrix Factorization algorithm is widely used in recommendation systems. However, its performance is often affected by data sparsity and high dimensionality of the data, and less consideration is given to the content information of items. In view of the above problems, we propose a Bipartite Network Representation learning regularized Matrix Factorization recommendation algorithm (BiNRMF). First, construct two bipartite networks using the rating information and the tag information of the item, respectively. Then, low-dimensional vectors of users and items are obtained by the Bipartite Network Embedding (BiNE) algorithm. These vectors can be applied to calculate the similarity between users and between items. Finally, traditional matrix factorization model is regularized by these similarities. Experimental results on the GoodBooks and MovieLens datasets demonstrate a significant improvement in prediction accuracy compared to the classical collaborative filtering and matrix factorization algorithms.

**Key words:** Recommendation Systems; Collaborative Filtering; Bipartite Network; Network Representation Learning; Matrix Factorization

# 1引言.

随着互联网的不断发展，数据呈几何式增长，信息爆炸时代来临。如何快速而精准地从浩瀚的数据海洋中帮助用户获取有效的信息，已成为亟需解决的问题。推荐系统作为一种有效的信息过滤技术，是解决这一问题的重要手段[1]。推荐系统的核心是对用户行为、物品属性和上下文等信息进行建模，推断出用户的兴趣爱好，并向用户推荐感兴趣的物品。目前,推荐系统已在很多领域得到了应用[2,3]，为企业带来了可观的商业价值。

在推荐系统领域，现有的推荐算法主要基于协同过滤思想[4-6]，包括基于记忆和基于模型的协同过滤方法。基于记忆的协同过滤算法的主要过程是：首先，通过计算相似度来寻找相似用户或相似物品；然后，使用相似用户或相似物品的评分值对缺失值进行预测。基于模型的协同过滤方法使用机器学习技术，通过历史数据训练一个模型，并基于模型来预测评分中的缺失值。虽然现在很多工作都在研究基于协同过滤的推荐方法，但依然存在一些重要的问题，例如数据的稀疏性使得一些样本之间可能不存在同维度的数据，无法获取可度量的相似度，导致相似度计算不稳定。数据的高维度导致相似度计算复杂度高，容易受噪声数据影响。以及如何融合外部信息以提高系统的推荐精度、解决冷启动等。最近的研究表明，相较于传统协同过滤算法中人工设计的相似度指标，使用网络表示学习的方法处理推荐系统中的相关信息，可以更有效地增强推荐系统的特征表示能力[7][8]。此外，推荐系统中天然存在着大量的网络结构，例如用户的评分信息、物品的标签信息都构成了二部网络。网络结构信息能为推荐算法提供丰富的输入，如何有效地利用网络结构数据是当前推荐系统研究的热点[9,10]。

在用户-项目二部网络中，虽然显式的连边关系只存在于不同类型的节点之间，但同一类型的节点之间存在隐式关系。例如，喜欢同一物品的两个用户之间就存在一种隐式的关系，这种关系表明用户之间存在相同的兴趣偏好。最近有论文指出，对这种隐式关系进行建模可以提高推荐的性能[11]。然而，现有大部分网络表示学习方法[12-14]仅对二部网络中节点的显式关系进行了建模，忽略了节点间潜在的隐式关系。为了能同时兼顾节点间的显式关系和隐式关系, 我们在学习节点表示时同时对显式关系和隐式关系进行建模。此外，在考虑用户的历史行为信息的同时，引入外部标签信息以缓解冷启动问题[15]。本文综合考虑用户-项目二部网络中的隐式关系和显式关系，并引入项目外部标签信息，主要贡献如下:

（1）针对用户-项目二部网络，使用二部网络表示学习的方法，学习用户的低维向量稠密表示，提高相似度计算结果的稳定性，降低相似度计算的复杂度；

（2）引入项目的标签信息构建物品-标签的二部网络，使用二部网络表示学习的方法，学习项目的低维向量稠密表示，从项目的内容角度来考虑物品之间的相似关系；

（3）将低维向量空间中用户与用户、项目与项目之间的相似关系融入传统矩阵分解算法，提出基于二部网络表示学习的矩阵分解推荐算法。

# 2 相关工作

## 2.1 矩阵分解模型

矩阵分解是实现协同过滤最常用的方法，矩阵分解模型因具有较好的理论基础、良好的扩展性等优点，受到了学术界和工业界的广泛关注。传统的矩阵分解模型[16]通过分解用户-物品的评分矩阵来获得用户和物品的潜在特征向量，然后根据两者的点乘结果预测用户对物品的喜好程度。Salakhutdinov 等人从概率角度解释了传统矩阵分解模型的合理性，并提出概率矩阵分解模型(probabilistic matrix factorization, PMF)[17]。在 Netflix 竞赛期间，Koren 等人[18]将邻域信息集成到矩阵分解中。它假定用户对特定物品的评分不仅由用户和该物品的潜在特征决定，而且还由用户对其他相似物品的评价行为构成，即考虑物品之间的邻域信息。这种方法在许多领域的性能都优于传统的矩阵分解模型。Wu 等人[19]通过标签信息来构造用户和物品的相似性关系，并将两者融入PMF中，提出基于近邻的概率矩阵分解模型。然而，现有的矩阵分解模型较少使用外部信息或者通过手工构造特征的方式使用外部信息，限制了外部信息的特征表达能力。

## 2.2 网络表示学习

近年来,网络表示学习逐渐成为机器学习中的一个热门的研究方向。网络表示学习试图为网络中的每个节点学得一个低维表示向量，并尽可能的保持其原有的结构信息[20]。DeepWalk [12]和 Node2vec [13]算法使用随机游走方法将网络转换为节点序列的语料库，并基于Word2Vec[21]模型学习节点的低维向量表示。此外，一部分工作考虑网络的高阶结构特性，例如 LINE [14]对节点中的一阶和二阶近邻关系进行建模。

但现有的网络表示学习方法主要针对同质网络[12,13]，Gao [20]指出这些方法应用于二部网络时忽略了二部网络的特殊性质，并提出二部网络表示学习算法BiNE [20]。BiNE同时考虑二部网络中不同类节点之间的显式关系和同类节点之间的隐式关系。通过联合优化这两个部分的目标函数来学习二部网络中节点的表示：

(1)



公式（1）中*O*1 代表对二部网络中的显式关系进行建模的目标函数，*O*2和*O*3分别代表对二部网络中用户与用户之间、项目与项目之间隐式关系进行建模的目标函数。参数分别代表了联合训练时各部分的比例。



最后,为了优化联合模型，使用随机梯度上升算法(SGA)对模型进行训练,获得网络中节点最终的低维向量表示。本文使用BiNE算法分别从用户-项目二部网络和项目-标签二部网络学习用户和项目的低维向量表示。

# 3 BiNRMF推荐算法

## 3.1 算法整体框架

本文结合二部网络表示学习算法和矩阵分解算法，利用两个异构信息源：物品的标签信息和用户的评分信息作为输入，通过训练模型预测用户对物品的评分。图1描述了算法的整体框架：首先，利用输入的信息分别构建用户行为和项目内容信息的二部网络。然后，使用二部网络表示学习的算法获得用户和项目的低维向量表示。最后，基于用户和项目的低维向量表示计算相似性，并融入矩阵分解模型，预测用户对项目的评分。

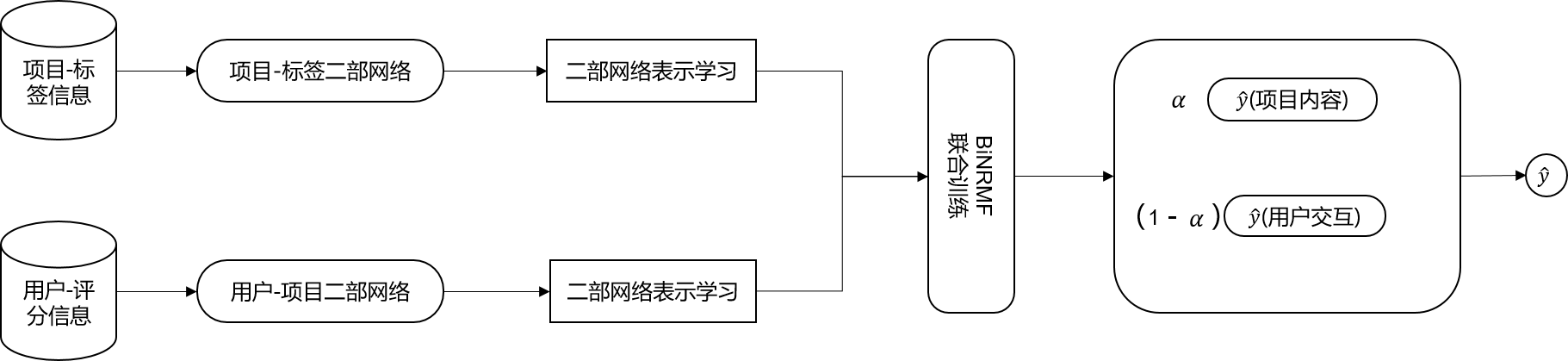


图1 算法整体框架

## 3.2 数据预处理

数据预处理的目的是从用户的评分行为和项目的内容信息中抽取用户-项目评分二部网络和项目-标签二部网络，并作为二部网络表示学习算法的输入。

如图 2 所示，左图为用户的评分信息，转化为右图的用户-项目评分二部网络。其中圆形节点代表用户，矩形节点代表项目。节点之间的边代表用户对物品存在打分行为。边的权重代表对应的用户项目评分。

如图3所示，左图为项目的标签信息，网络中的圆形节点代表的是项目，矩形节点代表项目标签，节点之间的边代表对应项目被用户打上对应标签的次数。

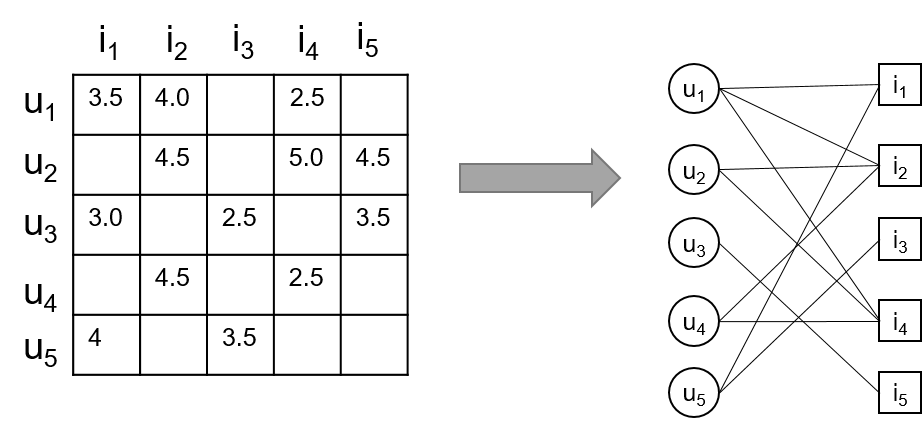


图2 用户评分信息二部图网络

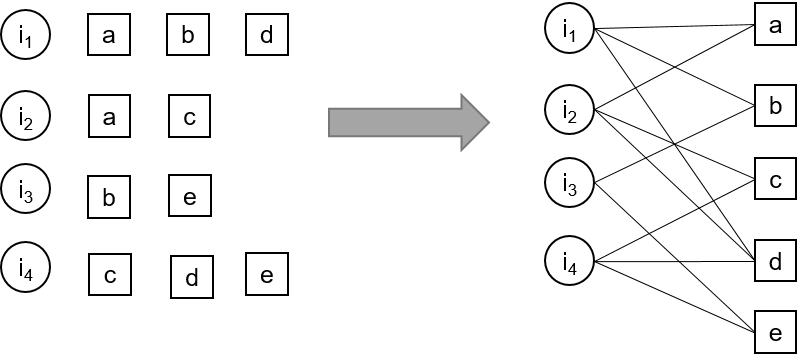


图3 项目标签信息二部图网络

最后，统一将二部网络的结构数据转化为二部网络表示学习算法所需要的格式，本文中数据的格式为：(节点A) \ （节点B）\ （边的权重）。

## 3.3 二部网络表示学习

获取到用户-项目二部网络和项目-标签二部网络的数据后，我们可将其应用到已有的二部网络表示学习算法。在本论文中,我们使用BiNE[20]算法分别对用户-项目二部网络和项目-标签二部网络进行表示学习，分别学得用户和项目的低维向量表示。

本文使用的两个网络均为加权网络。具体来说，用户-项目评分二部网络反映用户对项目的评分行为，其中连边的权重表示用户在对应项目上的评分。项目-标签二部网络反映了项目的内容信息，其中连边的权重表示标签在对应项目上被用户打上标记的次数。

## 3.4获取节点的近邻集合

通过表示学习的算法获取到用户和项目的低维向量表示后，我们可以在对应的低维表示空间中考虑用户之间和项目之间的相似性。本文使用余弦相似度来度量低维空间中向量的相似度。在计算网络中各个节点之间的相似性后，根据相似度阈值*θ*定义与目标节点相似的近邻节点集合。例如，*i* 用户的近邻节点集合定义为：

 （2）

节点之间的相似性信息将被用于正则化矩阵分解模型，而通过相似度阈值选取阈值范围内的近邻有助于减少相似度较低的节点对模型的干扰。

## 3.5 模型定义

矩阵分解是一种广泛采用的基于模型的协同过滤方法。传统的矩阵分解算法的目标函数如公式（3）所示:

(3)



该模型使用所有可以观测到的评分值来预测评分矩阵中的缺失值。然而，由于真实的评分矩阵非常稀疏，大多数情况下传统的矩阵分解不能生成最佳的预测值。因此，在本文中我们结合二部网络表示学习方法，提出了基于二部网络表示学习的矩阵分解推荐算法。同时建模用户之间的显式关系和隐式关系，并同时引入项目的外部内容信息，其目标函数定义为:

(4)



公式（4）中 代表物品*j*的相似近邻集合， 代表用户*i*的相似近邻集合。代表物品*j*和物品*k*在低维向量空间中的相似性，代表的是用户*i*和用户*k*在低维向量空间中的相似性。相似度的计算公式定义为：



(5)



在获得低维空间中节点的近邻集合后，我们基于协同过滤的思想，使用近邻集合中节点的评分来预测目标节点的评分，和表示分别使用基于项目的协同过滤算法和基于用户的协同过滤算法对传统矩阵分解模型进行修正，*α*代表了两个部分各自占的比例。

# 4实验分析

## 4.1 实验设置

为了衡量本文提出的 BiNRMF 推荐算法的性能，我们使用了MovieLens和GoodBooks两个数据集进行实验。MovieLens 数据集被广泛用于电影推荐系统[22]。该数据集包含600多用户对9000多电影的8万多条评分数据，该数据集是一个评分数据集，用户可以给电影评5个不同等级的分数（1-5分）。类似地，GoodBooks数据集中包含书籍和评分的详细信息，该数据集包含了5万多用户对1万本畅销书的评分数据。用户可以给书籍评5个不同等级的分数（1-5分）。为了使数据集保持一致，我们使用与MovieLens数据集相同的方式过滤GoodBooks数据集，仅保留数据集中至少有过100次交互的用户。表1总结了我们实验数据集的统计信息。

**表1：数据集的统计信息**

|  |  |  |
| --- | --- | --- |
|  | Goodbooks | MovieLens |
| |U| | 1167 | 610 |
| |V| | 10000 | 9742 |
| |E| | 129694 | 80419 |
| 连边密度 | 1.11% | 1.30% |

## 4.2 评估指标

在实验中，选择每个用户评分数据的80%作为训练集，剩下的20%作为测试集，保证每个用户都拥有一定量的评分数据。本文选择两个指标度量预测的精度，分别是绝对平均误差 (MAE)和均方根误差 (RMSE),定义如下:

(6)



 (7)

其中表示用户对项目的实际评分，表示用户对项目的预测评分，表示预测值的总数量。MAE 和 RMSE 是通过比较用户的真实评分和预测评分之间的差异来衡量推荐算法性能的好坏，它们的值越小，表示预测的精度越高。



## 4.3 性能评估

为评价所提出的 BiNRMF 方法的预测精度，选择以下经典方法在 MAE 和 RMSE 两个指标上进行对比实验：

* User-CF:该方法是基于用户的协同过滤算法，基于相似用户的评分预测缺失值。
* Item-CF:该方法是基于项目的协同过滤算法，基于相似项目的评分预测缺失值。
* MF:该方法是传统的矩阵分解算法，使用训练数据训练模型，基于模型预测评分值。
* NIMF[23]:该方法是融合邻域信息的矩阵分解算法。

各个方法的参数设置如表2所示:

**表2 算法重要参数**

|  |  |
| --- | --- |
| 方法 | 参数 |
| User-CF | 相似度阈值*θ*=0 |
| Item-CF | 相似度阈值*θ*=0 |
| MF  NIMF  BiNRMF | 矩阵维度*n*=10  矩阵维度*n*=10,融合比例*α*=0.8，邻居数量*n\_ neighbors*=50  矩阵维度*n*=10，融合比例*α*=0.8，相似度阈值*θ*=0 |

实验结果如表3所示。从表中可以看出，我们提出的方法MAE和RMSE明显小于其他方法的 MAE 和 RMSE。这表明通过使用二部图表示学习的方式，学到的用户和项目的低维向量表示,可以有效提高矩阵分解模型的预测性能。

**表3 性能对比**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方法 | MovieLens | | GoodBooks | |
|  | RMSE | MAE | RMSE | MAE |
| User-CF | 0.9060 | 0.7156 | 0.8585 | 0.6681 |
| Item-CF | 1.1158 | 0.8219 | 1.0272 | 0.8098 |
| MF | 1.4666 | 1.044 | 1.4454 | 1.0196 |
| NIMF | 0.8824 | 0.6807 | 0.8722 | 0.6923 |
| BiNRMF | **0.8563** | **0.6703** | **0.8448** | **0.6599** |

## 4.4 模型分析

(1) 参数α对实验结果的影响

在本文的方法中，参数*α*是决定 BiNRMF 方法依靠相似用户和相似项目的权重。若 *α*=1 时，仅使用基于用户的方法进行预测；若*α*=0 时，仅使用基于项目的方法进行预测。为研究*α*对预测结果的影响，本文在 MovieLens 和 GoodBooks 数据集上，设置矩阵维度*n*=20，阈值*θ*=0，*α*的值从 0.0 到 1.0并以 0.1 的间隔逐渐增加，观察对 MAE 和 RMSE 的影响，实验结果如图4所示。

****

图4 *α*值的影响

其中，图4(a)、图4(b)是两个数据集在RMSE的实验结果。图4(c)、图4(d)是在MAE上的实验结果。实验结果表明，参数*α*的值对 BiNRMF方法的预测精度有很重要的影响。

图4(a)中，在*α*=0.9 时，预测精度达到最优值；图4(c)中，在*α*=0.8 时,预测精度达到最优值。这表明，一个合适的*α*值可以达到更好的预测精度。同时，图4(a)、图4(c)的结果表明，在不同的数据集上，达到最优预测精度时的*α*值是不一样的。从实验结果我们可以看出，对于GoodBooks数据集和MovieLens数据集，项目内容的相似关系比用户行为的相似关系对预测结果影响更大。

(2)阈值*θ*对实验结果的影响

参数*θ*决定节点近邻集合的大小，平衡噪声和有用信息的比例。为研究阈值对预测精度的影响，实验分别在GoodBooks 和 MovieLens两个数据集上进行。实验设置*α*在两个数据集上分别取0.2和0.9，矩阵维度*n*=20,。设置*θ*值从 0 到 0.9 并以 0.1 的间隔逐渐增加，观察对 MAE 和 RMSE 的影响，结果如图5所示。



图5 阈值*θ*的影响

其中，图5(a)、图5(b)是*α*=0仅使用基于用户的方法进行预测，用户的相似度阈值对MAE影响的实验结果。图5(c)、图5(d)是*α*=1仅使用基于项目的方法进行预测时，项目的相似度阈值对MAE影响的实验结果。从图5可以看出，我们的模型对阈值的变化不敏感，模型具有很好的鲁棒性。

(3)矩阵维度*n*的影响

矩阵维度表示联合训练时矩阵的维度，影响最终的预测评分。为研究矩阵维度对预测精度的影响，实验分别在 GoodBooks 和 MovieLens 两个数据集上进行。实验设置*α*在两个数据集上分别取0.2和0.9，阈值*θ*=0，矩阵维度*n*的值从10到100，并以 10 的间隔逐渐增加，观察其对 MAE 和 RMSE 的影响。实验结果如图6所示。其中，图6(a)、图6(b)是两个数据集在RMSE的实验结果。图6(c)、图6(d)是在MAE上的实验结果.图6的实验结果表明，矩阵的维度对 NIMF 方法的预测精度有很重要的影响。从图6(a)、图6(c)中可以看出，一开始，随着矩阵维度的增大，预测精度一直在提高，当预测精度达到最大值之后，随着矩阵维度的增大，预测精度开始降低；图6中的四个实验，都在矩阵维度*n*=20 时，预测精度达到最优值。这表明矩阵维度过低时，隐藏向量空间对数据的刻画不够精细,导致预测精度下降，而维度过高时，可能会引入过多的噪声，同样会导致预测精度下降。因此，选择一个合适的矩阵维度*n*可以达到更好的预测精度。



图6 矩阵维度*n*的影响

# 5 结论与未来工作

本文提出一种基于二部网络表示学习的矩阵分解推荐算法BiNRMF。首先，根据用户评分历史信息和项目的标签信息，用二部网络表示算法 BiNE 分别对用户和物品进行表示；然后，在低维表示空间中获取用户和物品的近邻集合信息;最后，结合近邻集合信息和矩阵分解算法提出BiNRMF算法。我们在GoodBooks和MovieLens两个数据集的实验结果表明，与经典的协同过滤算法相比，该方法的预测精度有明显的提高。我们的工作假定融合协同过滤信息和项目内容信息所占的权重是固定的，但在实际场景下，针对不同的用户、不同的项目，我们的模型应该学习到不同的权重。后面的工作考虑采用注意力机制，使用深度学习的方式，自适应的融合协同过滤信息和项目内容信息在模型中所占的权重。

**参考文献**

[1] Burke R. Hybrid recommender systems: Survey and experiments[J]. User Modeling and User-Adapted Interaction, 2002, 12(4): 331-370.

[2] Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering[J]. IEEE Internet Computing, 2003 (1): 76-80.

[3] Zhang S, Yao L, Sun A, et al. Deep Learning Based Recommender System: A Survey and New Perspectives[J]. ACM Computing Surveys (CSUR), 2019, 52(1): 5.

[4] Miranda C, Jorge A M. Item-based and User-based Incremental Collaborative Filtering for Web Recommendations[C]//Portuguese Conference on Artificial Intelligence. Springer, Berlin, Heidelberg, 2009: 673-684.

[5] Sarwar B M, Karypis G, Konstan J A, et al. Item-based Collaborative Filtering Recommendation Algorithms[J]. WWW, 2001, 1: 285-295.

[6] Su X, Khoshgoftaar T M. A Survey of Collaborative Filtering Techniques[J]. Advances in Artificial Intelligence, 2009.

[7] Yang J, Li Z, Yin H, et al. Unified User and Item Representation Learning for Joint Recommendation in Social Network[C]//International Conference on Web Information Systems Engineering. Springer, Cham, 2018: 35-50.

[8] Wu H, Zhang H, He P, et al. A Hybrid Approach to Service Recommendation Based on Network Representation Learning[J]. IEEE Access, 2019, 7: 60242-60254.

[9] Tran T, Lee K, Liao Y, et al. Regularizing Matrix Factorization with User and Item Embeddings for Recommendation[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. ACM, 2018: 687-696.

[10] Zhang X, Zhao Z, Li C, et al. An Interpretable and Scalable Recommendation Method Based on Network Embedding[J]. IEEE Access, 2019, 7: 9384-9394.

[11] Yu L, Zhang C, Pei S, et al. Walkranker: A unified pairwise ranking model with multiple relations for item recommendation[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.

[12] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014: 701-710.

[13] Grover A, Leskovec J. node2vec: Scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2016: 855-864.

[14] Tang J, Qu M, Wang M, et al. Line: Large-scale information network embedding[C]//Proceedings of the 24th international conference on world wide web. International World Wide Web Conferences Steering Committee, 2015: 1067-1077.

[15] Shi S, Zhang M, Liu Y, et al. Attention-based Adaptive Model to Unify Warm and Cold Starts Recommendation[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. ACM, 2018: 127-136.

[16] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. Computer, 2009 (8): 30-37.

[17] Mnih A, Salakhutdinov R R. Probabilistic Matrix Factorization[C]//Advances in Neural Information Processing Systems. 2008: 1257-1264.

[18] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]//Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2008: 426-434.

[19] Wu L, Chen E, Liu Q, et al. Leveraging tagging for neighborhood-aware probabilistic matrix factorization[C]//Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012: 1854-1858.

[20] Gao M, Chen L, He X, et al. BiNE: Bipartite Network Embedding[C]//SIGIR. 2018: 715-724.

[21] Mikolov T, Sutskever I, Chen K, et al. Distributed Representations of Words and Phrases and Their Compositionality[C]//Advances in Neural Information Processing Systems. 2013: 3111-3119.

[22] He X, Gao M, Kan M Y, et al. Birank: Towards ranking on bipartite graphs[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 29(1): 57-71.

[23] Zheng Z, Ma H, Lyu M R, et al. Collaborative Web Service QoS Prediction via Neighborhood Integrated Matrix Factorization[J]. IEEE Transactions on Services Computing, 2012, 6(3): 289-299