

异质信息网络中的推荐算法研究

一、 研究背景

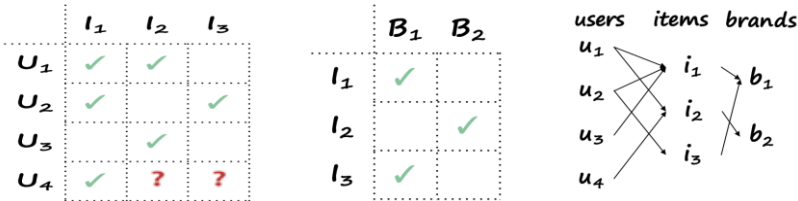
随大数据时代的来临，互联网信息爆炸式增长，而推荐是解决信息过载的有效方法，已广泛应用至电子商务和互联网服务中。作为重要的数据挖掘任务，推荐系统包括许多对象类型（如电影推荐中的用户、电影、演员和导演等）及对象间的丰富交互关系（如观看信息、社交信息和属性信息等）。因此，融合更多信息进行混合推荐是解决数据稀疏性的有效途径。

近年来，大量涌现的社会媒体网站包含许多类型对象的复杂交互，将这些对象建模为同质网络是很困难的。作为半结构化的表示方法，异质信息网络（Heterogeneous Information Network），即多种类型节点和链路的建模方式，可以有效建模和处理复杂多样的数据。其次，与同质网络相比，异质网络不仅可以融合更多类型的对象及其复杂交互关系，而且可以对齐多个社交网络平台的信息。最后，异质信息网络包含丰富的语义，将导致更细微的知识发现。因此，异质信息网络作为有效的信息融合方式可以用于建模推荐系统中的对象和关系，可能产生更准确的推荐结果。

二、 基础模型 NeuACF

经广泛调研，本文以发表于 IJCAI 2018 的 NeuACF（Neural network based Aspect-level Collaborative Filtering model）为基础，并加以改进。因此，本节具体介绍 NeuACF 的基本思想及模型结构。

虽然现有的隐因子模型在推荐领域表现出了强劲性能，但它们通常只挖掘购买历史的信息，而这仅从一个方面反映用户偏好和物品特征。但是，用户和物品的隐因子源于实际应用的不同方面。这些方面，从不同角度反映用户偏好和物品特征。以图 1 为例，进一步说明多方面信息的重要性。如果仅利用图 1（a）中的交互矩阵（即购买历史），会推断出用户 U_4 将购买物品 I_2 和 I_3 。然而，若考虑到图 1（b）所示的品牌信息，可以发现物品 I_3 可能更受用户 U_4 喜爱，因为物品 I_1 和 I_3 属于同一品牌 B_1 。



(a) 用户-物品交互矩阵 (b) 物品-品牌关系矩阵 (c) 异质信息网络

图 1 用户-物品方面级交互关系示例

尽管全面利用各方面信息是极有价值的，但面临两个严峻挑战。其一，如何提取不同方面级的信息。其二，如何从不同方面学习和融合隐因子。

为解决上述挑战，基于深度神经网络的 NeuACF 应运而生，可以有效提取和融合不同方面级的隐因子。具体地，推荐系统中不同类型的对象和其之间的交互被建模为异质信息网络，并利用元路径从不同方面提取用户和物品特征。如图 1 (c) 所示，可以通过用户-物品-用户路径，从购买历史角度提取用户的隐特征。此外，进一步利用深度神经网络捕捉复杂的非线性关系，为用户和物品学习不同方面级的隐因子，并基于注意力机制实现有效融合，进行 Top-N 推荐。

2.1 模型框架

NeuACF 模型包含三个主要步骤。首先，基于推荐系统中丰富的用户-物品交互信息构造异质信息网络，并在不同元路径下计算不同方面级的相似性矩阵，以反映用户和物品的不同方面级特征。接下来，利用深度神经网络，以所得相似性矩阵为输入，分别学习方面级隐因子。最后，结合注意力机制融合方面级隐因子，获得用户和物品的最终表示。接下来，详细说明以上三个步骤。

2.1.1 相似性矩阵计算

给定元路径，选择相似性矩阵提取方面级特征。原因有二，其一，相似性度量可以减弱噪声；其二，相似性数值在 $[0, 1]$ 之间，便于隐因子学习。因此，使用 PathSim 计算方面级相似性矩阵，如基于元路径 User-Item-Brand-Item-User (UIBIU) 和 Item-Brand-Item (IBI) 可以提取品牌方面特征。

2.1.2 方面级隐因子学习

已计算出不同方面级用户-用户和物品-物品相似性矩阵，接下来需要利用它们学习相应隐因子，架构如图 2 所示：

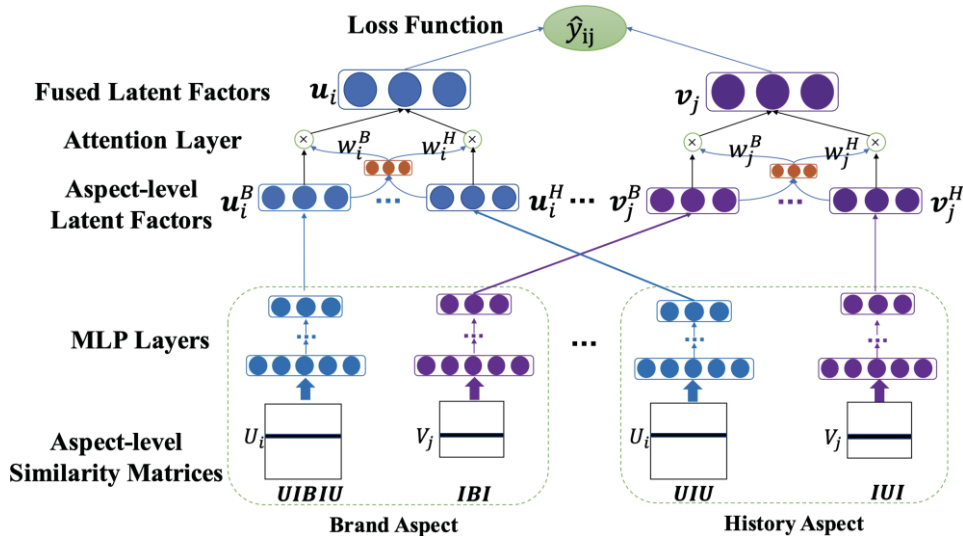


图 2 NeuACF 中的深度学习架构

具体地,以方面级相似性矩阵作为多层感知器(Multi-Layer Perceptron, MLP)的输入。相应地,方面级隐因子为其输出。以元路径UIBIU下的用户相似性矩阵 $S^B \in \mathbb{R}^{N \times N}$ 为例,用户 U_i 以 N 维的 S_{i*}^B 向量表示,该向量反映用户 U_i 与所有其他用户的相似性。其中, N 是用户总数。MLP 将用户 U_i 的初始相似性向量 S_{i*}^B 投影到低维方面级隐因子。形式化如下,给定初始输入向量 S_{i*}^B 和隐藏层 H_l ,其中 l 代表第 l 层,可以通过以下多层映射来学习最终的方面级隐因子 u_i^B :

$$\begin{aligned} H_0 &= S_{i*}^B \\ H_1 &= f(W_1^T * H_0 + b_1) \\ &\dots \\ H_l &= f(W_l^T * H_{l-1} + b_l) \\ &\dots \\ u_i^B &= f(W_n^T * H_{n-1} + b_n) \end{aligned}$$

其中, W_i 和 b_i 分别是第 i 层的权重矩阵和偏置, f 是隐藏层的激活函数。从图 2 中的架构可以看出,对于用户和物品的每个方面级相似性矩阵,均存在相应的 MLP 学习该方面级隐因子。

2.1.3 基于注意力机制的方面级隐因子融合

已获得用户和物品的方面级隐因子后,接下来需要将它们融合形成最终的表示。比较简单的方法是直接拼接所有方面级隐因子,或者平均所有方面级隐因子。但是,这两种方法都没有区分不同方面的重要性。这显然是不合理的,因为并非所有方面对于推荐都是有意义的。因此,考虑利用先进的注意力机制融合方面级隐因子。

注意力机制在多种机器学习任务中都表现出了优越的特性,如机器翻译。基于注意力机制,可以为所有方面级隐因子分配合适的权重:较高(较低)权重反映相应方面对于推荐是信息量充足(信息量匮乏)的。具体地,利用用户的品牌方面隐因子 u_i^B ,通过双层注意力网络计算相应权重:

$$S_i^B = W_2^T f(W_1^T * u_i^B + b_1) + b_2$$

其中, W 是权重矩阵, b 是偏置。

利用 Softmax 函数归一化,从而获得各方面级隐因子的最终权重,即方面 a 对用户 U_i 最终表示的贡献:

$$w_i^a = \frac{\exp(S_i^a)}{\sum_{n \in \mathbb{A}} \exp(S_i^n)}$$

其中， \mathbb{A} 是方面集合。

获得用户 U_i 的所有方面级隐因子的权重 w_i 后，可以计算最终表示 \mathbf{u}_i ：

$$\mathbf{u}_i = \sum_{a \in \mathbb{A}} w_i^a \cdot \mathbf{u}_i^a$$

2.2 模型优化

将 Top-N 推荐建模为分类问题，该问题预测用户与物品间的交互概率。为确保输出值是概率，需要将输出 \widehat{y}_{ij} 约束在 $[0,1]$ 内。使用 Logistic 函数作为输出层的激活函数，根据下式计算用户 U_i 和物品 I_j 间的交互概率：

$$\widehat{y}_{ij} = \text{sigmoid}(\mathbf{u}_i * \mathbf{v}_j) = \frac{1}{1 + e^{-\mathbf{u}_i * \mathbf{v}_j}}$$

其中， \mathbf{u}_i 和 \mathbf{v}_j 分别是用户 U_i 和物品 I_j 的最终表示。

在所有训练集上，利用似然函数的负对数形式，得到 point-wise 损失函数：

$$\text{Loss} = \sum_{i,j \in \gamma \cup \gamma^-} (y_{ij} \log \widehat{y}_{ij} + (1 - y_{ij}) \log(1 - \widehat{y}_{ij}))$$

其中， γ 和 γ^- 分别是正负实例集合， y_{ij} 是实例的 ground truth， \widehat{y}_{ij} 是预测结果。上式即为模型的目标函数，可以通过随机梯度下降或其变种技术来优化。

三、模型改进

3.1 NeuACF 问题分析

NeuACF 中，用户和物品的交互可能性是利用 embedding 间的点积度量的。由于点积本身所固有的缺陷——不满足三角不等式，可能导致相似性的传递特征被破坏，从而限制模型效果的提升。与此同时，NeuACF 使用 point-wise 损失函数，关注于评分信息的绝对数值。但对于 Top-N 推荐而言，相对交互可能性更为重要。接下来，本节具体分析以上两点问题可能对于模型效果产生的不良影响。



图 3 矩阵分解稳定解示意图

首先，如图 3 所示， U_1 、 U_2 和 U_3 是三个规模相同的用户组。其中， U_1 喜欢物品 v_1 ， U_2 喜欢物品 v_2 ， U_3 喜欢物品 v_1 和 v_2 。图 3 左半部分展示了在矩阵分解框架下的一组稳定解，即当用户喜欢物品时，用户和物品间的点积等于 2，否则点积等于 0。可以观察到，虽然用户 U_3 同时喜欢物品 v_1 和 v_2 ，但是物品 v_1 和 v_2 间的点积等于 0。这一点违反了三角不等式，即相似对 (U_3, v_1) 和 (U_3, v_2) 间的相似关系无法传播到 (v_1, v_2) 。

这种相似性传播被打破的情况，可能造成两种后果。第一，尽管矩阵分解框架在两个轴上捕获了最突出的隐因子，但是并没有反映用户 U_3 的反馈信息中包含的更细粒度偏好。第二，隐因子不能准确捕获用户-用户或物品-物品间的相似性，而这对于缓解协同过滤的数据稀疏性问题至关重要。

其次，point-wise 损失函数作为预测结果与 ground truth 的交叉熵，更为关注分数的绝对数值。而这种损失函数形式对于推荐任务而言，存在两点不足。第一，推荐产生的结果往往是一个物品列表，即 Top-N 推荐，因此追求的是排序的准确性，而非精确打分数值，只需相对打分即可。第二，point-wise 损失函数没有建模预测物品列表中的排序信息，可能会无意中过分强调那些不重要的物品，即排序结果后部，用户并不喜爱的物品。因此，Top-N 推荐效果可能受到限制。

3.2 改进方案

针对已分析的两点问题，本节阐述具体的改进方案及相应理论基础。

首先，相似性的传递与协同过滤技术密切相关。因为，在协同过滤中，也是通过已知的某些用户-物品对间关系，来推断其他未知关系的用户-物品对。正如上节所分析的那样，因三角不等式的性质未被满足，使用用户和物品间 embedding 的点积来捕捉已有的评分信息，并非是一种度量学习方式。从而，导致相似性的传递过程可能被打破。

为此，引入距离的概念。距离，是许多机器学习算法的核心概念，如 K-nearest neighbor (kNN)、K-means 等。度量学习利用距离定义的度量指标捕捉数据间的关系，从而使得相似对间距离较近，不相似对间距离较远。而度量学习所需满足的重要性质之一——三角不等式，自然地蕴含相似性的传递过程，即将已知的相似性信息通过距离传递给未知的点对，这与协同过滤的本质思想

不谋而合。因此，本文考虑以用户-物品度量指标替换 NeuACF 中的点积估计形式，在反映用户偏好的同时，捕捉用户-用户及物品-物品间的相似性，实现更全方位的信息传递。具体地，对于已知相似关系的用户-物品对集合 S ，度量指标将会使得 S 中相对应的 **embedding** 距离较近，而不在 S 中的点对距离较远。由于满足三角不等式，在此过程中会使得喜爱相同物品的用户聚成簇，且被相同用户喜爱的物品聚成簇。最终，对于用户的 kNN 物品将是：该用户曾喜爱的物品或具有相同口味的相似用户曾喜爱的物品。换句话说，通过学习满足已知相似信息的度量指标，不仅可以将这些信息传播到其他未知的用户-物品对，而且还可以传播给无法直接观察的用户-用户和物品-物品对。

形式化而言，将用户 i 的 **embedding** 记为 \mathbf{u}_i ，物品 j 的 **embedding** 记为 \mathbf{v}_j 。以它们之间的欧几里得距离作为度量指标，修改 NeuACF 打分函数：

$$d(i, j) = \|\mathbf{u}_i - \mathbf{v}_j\|$$

从而满足用户 i 对不同物品的相对喜爱程度，即喜爱的物品将比不喜爱的物品更接近该用户。

其次，既然已经在模型中引入了度量学习，那么相应地需要使用适合于度量学习的损失函数，即 **pair-wise** 损失函数。其输入是一个三元组 $\langle i, j, k \rangle$ ， i 是目标对象， j 是正样本，即同类对象， k 是负样本，即异类对象。对应于本文的推荐任务，也就是， i 是用户， j 是用户喜爱的物品， k 是用户不喜爱或不清楚喜好关系的物品。有如下不等式：

$$d(i, j)^2 + \text{threshold} < d(i, k)^2$$

其中， d 表示某种度量指标下的距离数值，**threshold** 表示特定阈值。该不等式本质上定义了正负样本之间的距离关系，即正样本间的距离与阈值之和会小于负样本间的距离。因此，基于此种三元组输入和不等式理论，相应的 **pair-wise** 损失函数为：

$$\mathcal{L}(d) = \sum_{(i, j) \in S} \sum_{(i, k) \notin S} [\text{threshold} + d(i, j)^2 - d(i, k)^2]_+$$

$$[z]_+ = \max(z, 0)$$

其中， S 是正样本集合。因此，最终的优化目标是拉近正样本 (i, j) 的距离，推远负样本 (i, k) 的距离。显然，只有括号内公式的值大于 0 时，才可以计算误差，并依据反向传播调节网络参数。

四、实验

4.1 实验设置

4.1.1 数据集

因时间因素限制，本文仅利用公开数据集Amazon评价模型性能。数据集的网络模式如图4所示，具体统计信息如表1所示。

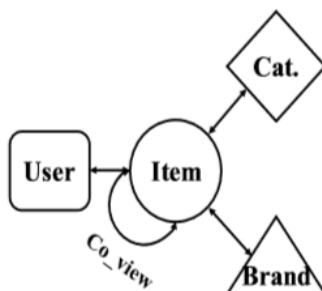


图4 Amazon数据集的网络模式

Datasets	#users	#items	#ratings	#density
Amazon	3532	3105	57,104	0.521%

表1 Amazon数据集的统计信息

Amazon: 该数据集包含亚马逊电商网站中的用户评分数据。在实验中，本文选择电子类商品进行模型评估。

公开数据集Amazon所包含的用户-物品交互信息为1至5的相应评分，因而直接建模会得到带权异质信息网络。但就本文所提模型而言，不考虑权重对于欧几里得距离度量指标的影响，因此将其统一处理为1。也就是说，若存在评分信息，相应用户-物品有连边；若不存在评分信息，则相应用户-物品无连边。

4.1.2 评价指标

本文采用leave-one-out法对模型性能进行评估。具体地，选择一个用户已评分的物品作为正样本，随机选择99个未被该用户评分的物品作为负样本。对于该用户而言，模型将对此100个物品进行排序，产生相应的推荐列表。为与baselines进行公平比较，本文对所有baselines测试集中每个（用户，物品）正样本，使用相同的负样本集合。

在此基础之上，通过命中率（Hit Ratio，HR）和归一化折损累计增益（Normalized Discounted Cumulative Gain，NDCG）作为具体评价指标。

与此同时，由于效率原因，本文并不产生全部的物品推荐列表，而是利用Top-K排序结果。其中， $K \in [5, 10, 15, 20]$ 。

4.1.3 Baselines

本文共利用8个基线进行对比实验，包括基础模型NeuACF、两个基本方法

(ItemPop和ItemKNN)、两个矩阵分解方法(Matrix Factorization, MF和element-wise Alternating Least Squares, eALS)、一个基于pair-wise损失函数的排序方法(Bayesian Personalized Ranking, BPR)和两个基于神经网络的方法(Deep Matrix Factorization Models, DMF和Neural networkbased Collaborative Filtering, NCF), 概要介绍如下:

NeuACF, 将推荐系统建模为异质信息网络, 利用元路径捕捉用户和物品不同方面级隐因子, 并实现高效融合。

ItemPop, 物品按照受欢迎程度, 即与所有用户的交互总次数, 排序推荐。

ItemKNN, 基于物品的标准协同过滤方法。

MF, 矩阵分解框架中最具有代表性的隐因子模型。

eALS, 用于推荐的矩阵分解算法, 基于element-wise交替最小二乘技术, 有效优化具有可变加权缺失数据的矩阵分解模型。

BPR, 利用贝叶斯后验估计, 实现个性化排序, 并进一步使用pair-wise损失函数优化矩阵分解模型。

DMF, 利用深度神经网络的矩阵分解模型, 将用户和物品映射到共同的低维向量空间。

NCF, 基于神经网络的协同过滤算法, 结合了矩阵分解的线性和深度神经网络的非线性, 用于对用户-物品隐结构进行建模。

4.1.4 实现细节

本文基于TensorFlow深度学习框架实现所提模型, 并对所有数据集, 调节相应超参数, 使得模型达到最佳效果。对于神经网络, MLP层数在集合{2, 3, 4}中调节, 每个隐藏层具有600个隐藏单元。潜在因子维数在集合{16, 32, 64, 128}中调节。batch大小在集合{512, 1024}中调节, 学习率在集合{0.00005, 0.0001}中调节。在训练模型时, 对于每个正样本, 负采样的负样本数量在集合{2, 10, 12}中调节, 且度量中的margin在集合{0.5, 1, 2}中调节。模型用Xavier初始化器随机初始化模型参数, 并使用Adam作为优化器优化模型。表2说明了实验中所用的元路径及相应方面。与此同时, 对于所有baselines, 根据相应文献, 设定其最佳参数。

数据集	方面	元路径	
		用户	物品
Amazon	历史	UIU	IUI
	品牌	UIBIU	IBI
	种类	UICIU	ICI
	共同浏览	UIVIU	IVI

表2 实验所用元路径及其对应方面

4.2 实验结果

Datasets	Metrics	ItemPop	ItemKNN	MF	eALS	BPR	DMF	NCF	NeuACF	Ours
Amazon	HR@5	0.2412	0.1897	0.3027	0.3063	0.3296	0.2693	0.3117	0.3268	0.3595
	NDCG@5	0.1642	0.1279	0.2068	0.2049	0.2254	0.1848	0.2141	0.2232	0.2319
	HR@10	0.3576	0.3126	0.4278	0.4287	0.4657	0.3715	0.4309	0.4686	0.5313
	NDCG@10	0.2016	0.1672	0.2471	0.2441	0.2693	0.2179	0.2524	0.2683	0.2873
	HR@15	0.4408	0.3901	0.5054	0.5065	0.5467	0.4328	0.5258	0.5591	0.6554
	NDCG@15	0.2236	0.1877	0.2676	0.2647	0.2908	0.2332	0.2774	0.2924	0.3202
	HR@20	0.4997	0.4431	0.5680	0.5702	0.6141	0.4850	0.5897	0.6257	0.7370
	NDCG@20	0.2375	0.2002	0.2824	0.2797	0.3067	0.2458	0.2925	0.3080	0.3396

表 3 不同方法的 HR@K 和 NDCG@K 指标对比

表 3 展示了不同方法的实验结果。其中，峰值性能用粗体标出。可以看出，本文所提模型在 Amazon 数据集的所有指标上都达到了峰值性能。具体地，Amazon 数据集较稀疏，在这种情况下，本文所引入的基于欧几里得距离的度量指标凸显其相应作用，即不仅刻画了显式的用户-物品间关系，也捕捉了隐含的更细粒度的用户-用户和物品-物品相似性，从而有效缓解数据稀疏问题。与此同时，也进一步说明了 pair-wise 损失函数相比 point-wise 损失函数，显式约束了排序信息，可以得到排序更为合理的推荐列表。

观察 baselines 表现，可以看到，NeuACF 在 Amazon 数据集的多数指标上都能达到峰值性能，说明 NeuACF 作为基于异质信息网络的最新模型，在推荐领域仍具有不俗的表现。究其原因，通过利用异质信息网络这一强有力工具，NeuACF 基于不同元路径学习到不同方面级隐因子，从多角度提供了用户和物品的特征，从而得到全面的表示学习结果。

五、总结

本文通过研究背景介绍，从宏观角度阐述了基于异质信息网络进行推荐的原因。进一步地，深入分析了基础模型 NeuACF 的具体结构，并论述了其存在的两点问题，给出相应改进方案。具体地，引入 metric learning 和 pair-wise 损失函数，产生排序更合理的推荐列表。最终通过编码实现改进模型，在推荐数据集 Amazon 上，与经典算法进行了对比实验，验证了所提模型的优越性。

综上，本文取得的主要成果有三点：

- 1、在基础模型 NeuACF 中引入度量学习的概念，以欧几里得距离指标捕捉数据间关系，弥补点积所造成的相似性传递特征丢失，实现更全方位的信息传递。
- 2、使用 pair-wise 损失函数替换 point-wise 损失函数，强调对于 Top-N 推荐而言的相对交互可能性，产生排序更合理的物品推荐列表。
- 3、基于推荐数据集 Amazon，与经典算法进行对比实验，验证了模型的优越

性。

六、 问题和展望

本文实现的改进模型在稀疏数据集 Amazon 上表现优良。但是，对于更稠密的数据集，可能存在收敛速度变缓且性能降低的问题。如何平衡性能和数据数量，是未来努力的方向。与此同时，在实际应用中，不仅存在用户评分这样的显式反馈信息，而且充斥着大量诸如点击、加入购物车等隐式反馈，这些信息对于刻画用户喜好同样起着关键作用。本文由于时间有限，未搜寻合适的隐式反馈数据集评估模型，这同样也是未来的工作之一。