Heterogeneous Information Network Embedding for Recommendation

# Abstract

由于数据异构性建模的灵活性，异构信息网络（HIN）被用来描述推荐系统中复杂和异构的辅助数据，称为基于HIN的推荐。从HIN提取和利用信息开发基于HIN的推荐的有效方法是具有挑战性的。**大多数基于HIN的推荐方法依赖于基于路径的相似度**，不能充分挖掘用户和项目的潜在结构特征。在本文中，我们**提出了一种新颖的基于HIN的推荐的异构网络嵌入方法**，称为HERec。为了嵌入HIN，我们设计了一个基于元路径的随机游走策略，**为网络嵌入生成有意义的节点序列**。**学习节点嵌入首先被一组融合函数转换，然后被整合到扩展矩阵分解（MF）模型中**。**将扩展的MF模型与融合函数一起针对评估预测任务进行联合优化。**在三个真实世界的数据集上的大量实验证明了HERec模型的有效性。此外，我们还展示了HERec模型在冷启动问题上的能力，揭示了HINs的变换嵌入信息可以提高推荐性能。

# INTRODUCTION

作为一种有前途的方向，由多种节点和链路组成的异构信息网络（HIN）已被提出作为一种强大的信息建模方法[4] - [6]。由于其在建模数据异构性方面的灵活性，在推荐系统中已经采用HIN来表征丰富的辅助数据。在图1中，如图1（a）所示，我们给出一个以HIN为特征的电影推荐的例子。我们可以看到，HIN包含了不同类型的关系连接的多种类型的实体。**在基于HIN的表示下，推荐问题可以被认为是HIN上的相似搜索任务**[4]。这种推荐设置被称为基于HIN的推荐[7]。基于HIN的推荐在文献[7] - [12]中受到了很多关注。大多数现有的基于HIN的推荐方法的基本思想是利用HIN上的用户和项目之间的基于路径的语义相关性，例如基于元路径的相似度来进行推荐。

虽然基于HIN的方法已经在一定程度上实现了性能改进，但是使用基于元路径的相似性的这些方法存在**两个主要问题**。首先，基于元路径的相似性**依赖于显式的路径可达性**，并且**当路径连接稀疏或有噪声时，可能不可靠用于推荐**。**HIN中的一些链接可能偶然形成**，不能传达有意义的语义。其次，基于元路径的相似性主要表征HIN上定义的语义关系，不一定直接适用于推荐系统。在某些情况下，**基于相似性的派生路径很可能对推荐性能没有明确的影响**。现有方法主要是学习一种线性加权机制来**结合基于相似性的路径[11]或结合潜在因素**[7]，但是他们**不能学习HIN信息的复杂映射机制**进行推荐。这两个问题基本反映了基于HIN的推荐的两个基本问题，即基于HIN推荐的有效的信息提取和利用。

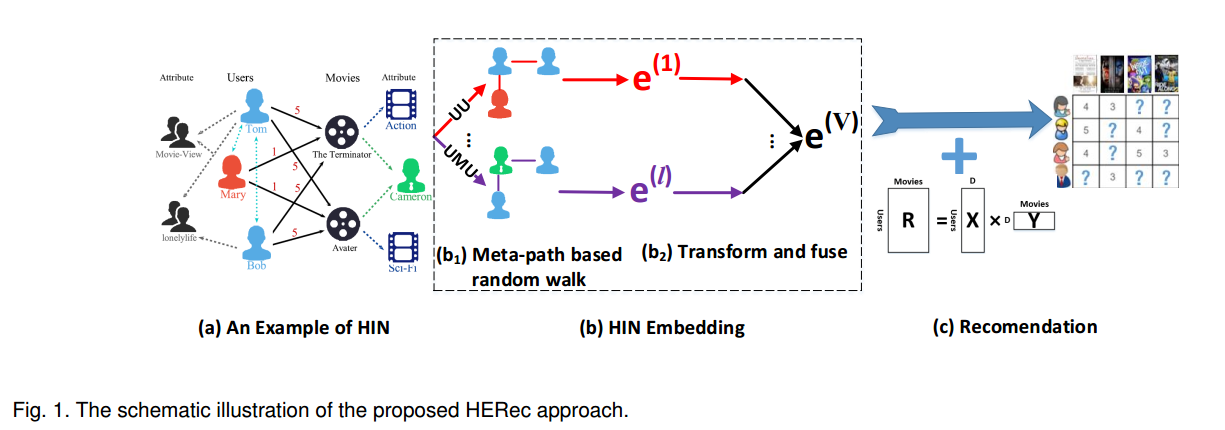
对于第一个问题，由于数据的异质性，开发一种有效提取和表示HIN有用信息的方法是具有挑战性的。与以往使用基于元路径的相似性的研究[4]，[7]不同，我们的想法是学习有效的异构网络表示来总结HIN的重要结构特征和属性。**在[13]，[14]之后，我们使用低维向量来表征来自HIN的节点，即嵌入**。我们不想依赖明确的路径连接，而是使用潜在的向量来编码有用的信息。与基于元路径的相似度相比，**学习的嵌入更加紧凑**，易于使用和集成。另外，网络嵌入方法本身对于稀疏和嘈杂的数据更有抵抗力。然而现有的大多数网络嵌入方法都只关注同构网络，只能由单一类型的节点和链路组成，不能直接处理由多种节点和链路组成的异构网络。因此，我们提出了一种新的异构网络嵌入方法。考虑到元路径所反映的异构特征和丰富的语义，本文提出的方法首先采用由元路径引导的随机游走策略来生成节点序列。对于每一个元路径，我们通过最大化与根据给定的元路径采样的序列中的相邻节点的共现概率，来学习节点的唯一嵌入表示。我们融合了多个嵌入w.r.t. 与HIN嵌入的输出不同的元路径。

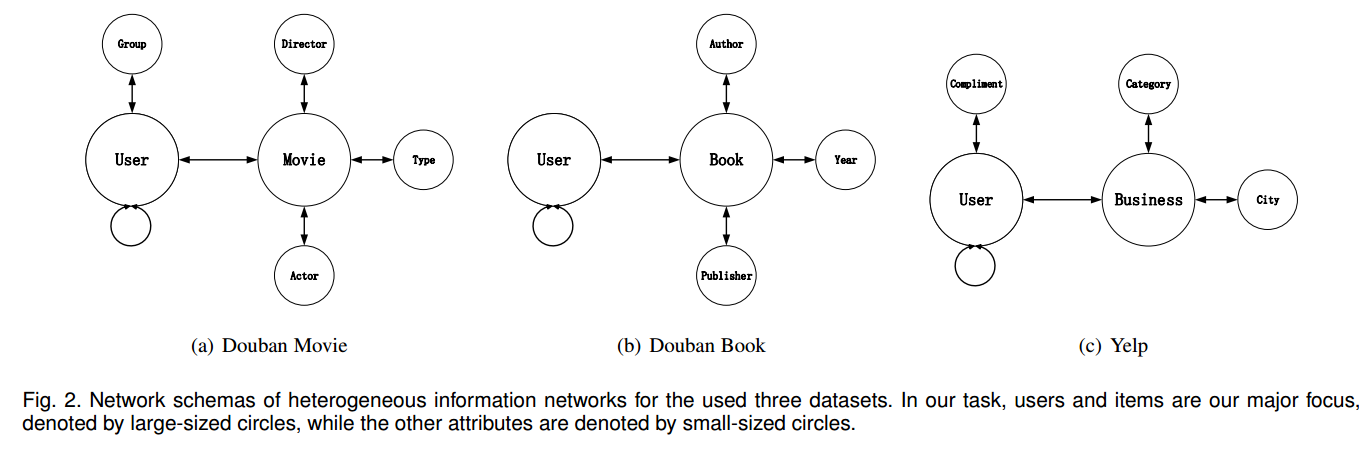
在获得HIN的嵌入后，我们研究如何在推荐系统中整合和利用这些信息。我们不假定学习嵌入自然适用于推荐系统。相反，我们提出并探索三个融合函数，将一个节点的多个嵌入融合成一个单一的表示进行推荐，包括简单线性融合，个性化线性融合和非线性融合。这些融合函数提供了灵活的方法将HIN嵌入转换为有用的信息进行推荐。特别是，我们强调个性化和非线性是我们设定的信息转换的两个关键点。最后，通过融合融合的HIN嵌入来扩展经典的矩阵分解框架。对预测模型和融合函数进行联合优化，进行评分预测任务。

# 2RELATED WORK

在本节中，我们将从推荐系统，异构信息三个方面回顾相关研究网络和网络嵌入。介绍的非常详尽。

# 3 PRELIMINARY

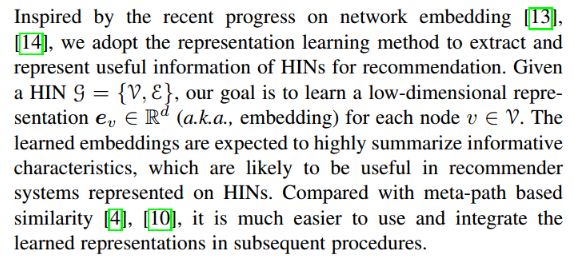




# 4 THE PROPOSED APPROACH

在本节中，我们提出了一种称为HERec的基于异构网络的*Rec*emendation嵌入方法。为了解决第一节介绍的两个问题，所提出的HERec方法由两个主要部分组成。首先，我们提出了一种新的异构网络嵌入方法来学习HIN中的用户/项目嵌入。然后，我们通过使用灵活的融合函数集合学习的嵌入来扩展经典的矩阵分解框架。我们给出了图1中提出的方法的总体示意图。HINs的建立（图1（a）），提出了两个主要步骤，即HIN嵌入（图1（b））和推荐（图1（c））。接下来，我们将详细介绍所提出的方法。

## 4.1 Heterogeneous Network Embedding



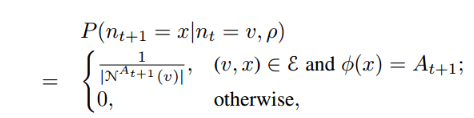
预计**学习到的嵌入**将高度概括信息特征，这对于在HIN上表示的推荐系统可能是有用的。与基于元路径的相似性[4]，[10]相比，在以后的过程中使用和整合学习的表示要容易得多。

然而，现有的网络嵌入方法大多集中在同构网络上，不能有效地对异构网络进行建模。例如，deepwalk的开创性研究[13]使用随机游走来生成节点序列，不能区分不同类型的节点和边缘。因此，它需要一个更原则的方法来遍历HIN并生成有意义的节点序列。

### 4.1.1 Meta-path based Random Walk

为了生成有意义的节点序列，关键是要设计一个有效的步行策略，能够捕捉HIN中反映的复杂语义。在HIN的文献中，元路径是描述HIN语义模式特征的重要概念[4]。因此，我们建议使用基于元路径的随机游走方法来生成节点序列。

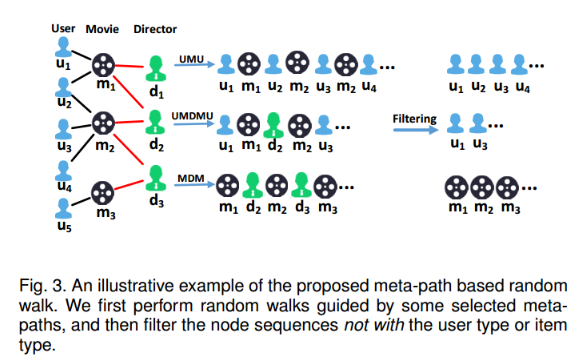
the walk path is generated according to the following distribution：



直观地看到这些元路径对应于不同语义关系可以生成有意义的节点序列。

### 4.1.2 Type Constraint and Filtering

因为我们的目标是提高推荐性能，所以主要的重点是学习用户和项目的有效表示，而其他类型的对象对我们的任务不太感兴趣。因此，我们**只选择以用户类型或项目类型开始的元路径**。一旦使用上述方法生成节点序列，就可能包含不同类型的节点。**我们进一步删除与起始类型不同类型的节点**。这样，最后的序列将只包含具有起始类型的节点。**对节点序列应用类型过滤有两个好处**。首先，虽然节点序列是使用具有异质类型的元路径构建的，但**最终的表示是使用同质邻居**（homogeneous neighborhood）**学习的**。我们在相同的空间中嵌入相同类型的节点，这放松了在统一空间中表示所有异构对象的具有挑战性的目标。其次，**给定一个固定长度的窗口，节点能够利用更多的同质邻居，这些邻居比不同类型的邻居更有可能是相关的**。

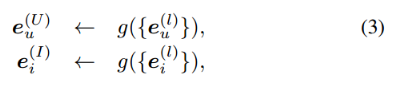


### 4.1.3 Optimization Objective

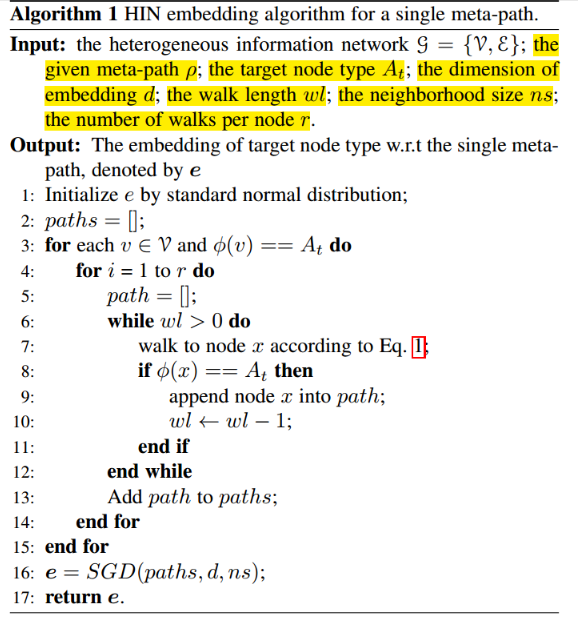
### 4.1.4 Embedding Fusion

**多种元路径表示的融合问题**

异构网络嵌入提供了从HIN中提取有用信息的一般方法。对于我们的模型，给定一个节点v，我们可以得到一组表达式，其中P表示元路径集。它需要一个原则性的融合方式来将节点嵌入转换成更适合的形式，这对于提高推荐性能是有用的。现有的研究通常**采用线性加权机制来组合从HIN挖掘出的信息**（例如，基于元路径的相似性），这可能无法获得用于推荐的有效信息表示。因此，我们建议使用一个通用函数g（·），其目的是融合用户和项目的学习节点嵌入：



由于用户和项目是我们的重点，我们只学习用户和项目的嵌入。在现阶段，我们没有指定函数g（·）的形式。相反，我们认为要根据具体的任务来学习好的融合功能。因此，我们在我们的推荐模型中留下了融合函数的表达和优化。



## 4.2 Integrating Matrix Factorization with Fused HIN Embedding for Recommendation

### 4.2.2 Setting the Fusion Function

**简单的线性融合。**我们假设每个用户对每个元路径具有相同的偏好，因此，我们为每个用户分配具有统一权重（即平均值）的元路径。此外，我们线性转换嵌入到目标空间。

**个性化的线性融合。**简单的线性融合不能模拟用户对元路径的个性化偏好。因此，我们进一步为每个用户分配元路径上的权重矢量，表示用户对每个元路径的个性化偏好。每个用户在许多实际应用中都有他/她的个人兴趣偏好是更合理的。

**个性化的非线性融合。**线性融合在复杂的数据关系建模中表现力有限。因此，我们使用非线性函数来提高融合能力。