

面向 HIN 基于多视角嵌入融合的推荐方法*

石乐昊¹, 寇月¹, 申德荣¹, 聂铁铮¹, 李冬²



¹(东北大学 计算机科学与工程学院, 辽宁 沈阳 110169)

²(辽宁大学 信息学院, 辽宁 沈阳 110036)

通信作者: 寇月, E-mail: kouyue@cse.neu.edu.cn

摘要: 由于异构信息网络 HIN (heterogeneous information network) 具有丰富的语义信息而在推荐任务中得到广泛应用. 传统的面向异构信息网络的推荐方法忽略了网络中关联关系的异质性, 以及不同关联类型之间的相互影响. 提出了一种基于多视角嵌入融合的推荐模型, 分别从同质关联视角和异质关联视角来挖掘异构信息网络的深层潜在特征并加以融合, 有效地保证了推荐结果的准确性. 针对同质关联视角, 提出了一种基于图卷积神经网络的嵌入融合方法, 通过对同质关联作用下节点邻域信息的轻量式卷积, 实现节点嵌入的局部融合. 针对异质关联视角, 提出了一种基于注意力的嵌入融合方法, 利用注意力机制来区分不同关联类型对节点嵌入的影响, 实现节点嵌入的全局融合. 通过实验验证了所提出的关键技术的可行性和有效性.

关键词: 异构信息网络; 推荐; 多视角嵌入融合; 图卷积神经网络; 注意力机制

中图法分类号: TP18

中文引用格式: 石乐昊, 寇月, 申德荣, 聂铁铮, 李冬. 面向 HIN 基于多视角嵌入融合的推荐方法. 软件学报, 2022, 33(10): 3619–3634. <http://www.jos.org.cn/1000-9825/6632.htm>

英文引用格式: Shi LH, Kou Y, Shen DR, Nie TZ, Li D. Recommendation Method Based on Multi-view Embedding Fusion for HINs. Ruan Jian Xue Bao/Journal of Software, 2022, 33(10): 3619–3634 (in Chinese). <http://www.jos.org.cn/1000-9825/6632.htm>

Recommendation Method Based on Multi-view Embedding Fusion for HINs

SHI Le-Hao¹, KOU Yue¹, SHEN De-Rong¹, NIE Tie-Zheng¹, LI Dong²

¹(School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China)

²(College of Information, Liaoning University, Shenyang 110036, China)

Abstract: HINs (heterogeneous information networks) have rich semantic information, which are widely used in recommendation tasks. Traditional recommendation methods for heterogeneous information networks ignore the heterogeneity of association relationships and the interaction between different association types. In this study, a recommendation model based on multi-view embedding fusion is proposed, which can effectively guarantee the accuracy of recommendation by mining the deep potential features of networks from the view of homogenous association and heterogeneous association respectively. For the view of homogenous association, a graph convolutional network (GCN)-based embedding fusion method is proposed. The local fusion of node embeddings is realized through the lightweight convolution of neighborhood information under the action of homogeneous associations. For the view of heterogeneous association, an attention-based embedding fusion method is proposed, which uses attention mechanism to distinguish the influence of different association types on node embedding, and realizes the global fusion of node embedding. The feasibility and effectiveness of the key technology proposed in this study are verified by experiments.

Key words: heterogeneous information network (HIN); recommendation; multi-view embedding fusion; graph convolutional network (GCN); attention mechanism

* 基金项目: 国家自然科学基金(62072084, 62172082, 62072086, U1811261); 辽宁省教育厅科学研究项目(LJKZ0094)

本文由“智慧信息系统新技术”专题特约编辑邢春晓研究员、王鑫教授、张勇副研究员、于戈教授推荐.

收稿时间: 2021-07-20; 修改时间: 2021-08-30; 采用时间: 2021-12-24; jos 在线出版时间: 2022-02-22

现实世界中存在大量的对象, 这些对象及其相互之间的关联关系形成了一个庞大而复杂的网络, 即信息网络. 随着对象类别及关联关系的复杂化, 信息网络逐渐向异构化方向发展, 异构信息网络包含多种类别的对象和关联关系. 由于异构信息网络具有丰富的语义信息, 因而被越来越多地应用于推荐系统, 处理复杂的多元异构的数据. 在本文主要讨论的电商场景下的商品推荐问题中, 用户与项目(商品)之间会产生“浏览”“收藏”“购买”等各种类型的交互数据, 这些数据所形成的异构信息网络结构优良, 蕴含了大量信息, 可以用于为用户提供高质量的推荐(如图 1 所示). 因此, 如何充分利用异构信息网络异构、多样化的特点, 如何深入挖掘网络中所隐含的信息并应用于推荐系统中, 是需要亟待研究的问题.

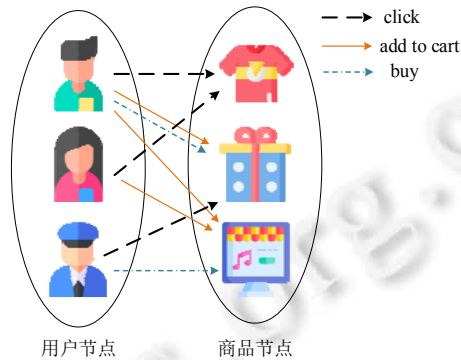


图 1 面向异构信息网络的推荐示例

利用网络嵌入技术可以对异构信息网络进行表征学习, 用低维、稠密的向量空间表示高维、稀疏的向量空间, 所学习到的特征可以用于计算用户与项目之间的相关性, 从而确定推荐目标. 目前, 很多工作都围绕网络嵌入技术展开了研究. 例如, 从文献[1]提出的 DeepWalk 模型到文献[2]提出的 node2vec 模型, 这些技术旨在网络中使用随机游走获取节点的邻域信息, 基于邻域信息对该节点进行嵌入. 文献[3]在 Deepwalk 的基础上提出了一种元路径的网络嵌入方法 metaPath2vec, 通过人为定义元路径来提高训练效果, 提高了语义相关性的识别能力. 近年来, 研究者更为关注如何使用图卷积神经网络(graph convolutional network, GCN)进行网络嵌入并应用于推荐系统之中. 例如, 文献[4]设计了采样技术进行图卷积计算, 以减少推荐过程中的计算难度. 文献[5]针对图卷积网络的协同过滤进行设计, 迭代传播邻域节点信息, 为有价值的邻域信息进行潜层空间的建模.

然而, 当前的网络嵌入技术大多只考虑单一视角下的网络特征(如: 仅考虑单一类型的关联关系), 忽略了异构信息网络中关联关系的异质性. 虽然一些研究工作考虑了多个视角下的网络特征及关联关系的多样性, 但只是简单地将这些特征进行整合, 缺乏对交互行为的深层表示和挖掘, 也忽略了不同关联类型之间的相互影响, 难以满足推荐系统的实际需求. 如图 1 所示, 用户和商品被划分为了不同的节点类型, 而用户与商品之间也产生多种行为关系. 传统的节点嵌入技术可以聚合同类行为下邻域节点与当前节点之间的关系, 但忽视了不同交互行为之间的关系. 例如, 用户 A 和用户 B 都购买了商品 C, 用户 B 点击了商品 D 又把商品 E 加入了购物车. 传统的基于网络嵌入的推荐模型往往忽略了多种行为之间的相互影响, 无法有效挖掘和利用高阶行为之间的交互关系.

当前, 面向异构信息网络的推荐模型的研究面临如下挑战: 首先, 异构信息网络中存在复杂的节点类型及关联类型, 如何将这复杂的信息应用于推荐系统之中? 例如, 两个用户存在相似的商品浏览记录, 基于协同过滤思想, 可以为其中一位用户推荐另一位用户的历史购买清单. 然而, “浏览”和“购买”是两种不同的用户行为, 如果不加区分, 将影响推荐结果的准确性. 其次, 如何区分不同关联类型对于节点嵌入所产生的贡献? 虽然, 通过引入图注意力机制可以建模上下文对于节点产生的贡献, 细化了模型的表达, 但也使得模型参数过于庞大, 而往往只能构建较为简单的模型, 难以训练异构信息网络这样复杂的网络结构.

针对上述问题, 本文面向异构信息网络提出了一种基于多视角嵌入融合的推荐方法, 主要贡献包括:

(1) 提出了一种基于多视角嵌入融合的推荐模型(recommendation model based on multi-view embedding fusion, MVEF). 与传统推荐模型不同, 该模型考虑了异构信息网络中关联关系的异质性, 分别从同质关联视角和异质关联视角来挖掘异构信息网络的深层潜在特征并加以融合, 充分利用了网络中所蕴含的丰富语义信息, 有效地保证了推荐结果的准确性.

(2) 针对同质关联视角, 提出了一种基于图卷积神经网络(graph convolutional network, GCN)的嵌入融合方法, 通过对同质关联作用下节点邻域信息的轻量式卷积, 实现节点嵌入的局部融合.

(3) 针对异质关联视角, 提出了一种基于注意力的嵌入融合方法, 利用注意力机制来区分不同关联类型对节点嵌入的影响, 实现节点嵌入的全局融合.

(4) 通过实验验证了本文所提出的关键技术的可行性和有效性.

本文第 1 节介绍异构信息网络推荐的相关工作. 第 2 节对本文的研究问题进行定义. 第 3 节提出基于多视角嵌入融合的推荐模型 MVEF. 第 4 节提出基于 MVEF 的推荐算法. 第 5 节为实验部分, 对提出的模型及算法进行测试. 最后对全文进行总结, 并给出下一步研究计划.

1 相关工作

本节主要介绍并讨论用于异构信息网络的推荐算法的相关工作, 包括: 基于图卷积神经网络的推荐模型、基于图注意力机制的表示模型和基于元信息(包括元路径和元图)的推荐模型. 然后, 将本文提出的关键技术与这些技术进行比较.

1.1 基于GCN的推荐模型

近年来, 图卷积神经网络由于其强大的结构化数据表示学习能力而取得了巨大的成功. 最近, 这类方法在推荐中引起了越来越多的关注. 例如, 文献[6]提出 GC-MC 模型将图卷积网络应用于用户-项目图, 它使用一个卷积层来扩展用户和项目之间的直接连接. 文献[4]将随机游走与项目图上的多个图卷积层相结合, 用于图像推荐. 文献[5]通过在用户项交互图上传播嵌入来利用高阶近似. 文献[7,8]进一步提出 LR-GCCF 和 LightGCN 模型, 它们通过去除嵌入传播层中的非线性激活函数和特征变换来提高协同过滤推荐任务的性能.

1.2 基于图注意力机制的表示模型

对于用 GCN 构建节点嵌入的方法所存在的问题, 2018 年, 文献[9]提出 GAT (graph attention network), 将注意力机制应用于图结构的数据中, 为不同的邻居分配不同的权重, 该方法引入了注意力在图模型中的基本使用机制. GAT 基于邻域消息传播模型, 将异构节点投影到统一的向量空间中, 然后通过 softmax 为邻居节点计算注意力系数 α_{ij} , 用注意力系数取代 MLP 中的线性参数矩阵, 从而得到了基本的注意力结构.

2019 年, 在 GAT 的基础上, 文献[10]提出 GATNE, 将应用场景扩展到了多重属性异构网络(attributed multiplex heterogeneous network, 即本文中定义的异构信息网络)中. GATNE 按照无向图和有向图分别计算各类边的嵌入, 在每个嵌入的计算中使用注意力机制连接邻居节点, 使其可以用于推导式任务, 生成的模型可以应用于未训练过的网络. 2020 年, 文献[11]对于异构信息网络问题上的应用, 基于元路径内部和元路径间聚合的方式进行建模, 并提出了加入多头的注意力机制, 以稳定模型的效果.

1.3 基于元信息的推荐模型

异质信息网络(heterogeneous information network, HIN)作为一种新兴的方向, 可以很自然地推荐系统中复杂的对象及其丰富的关系进行建模, 其中, 对象的类型不同, 则对象之间的连接代表不同的关系. 在此基础上, 许多研究者纷纷提出了多种基于路径的相似性度量方法, 用于评价异质信息网络中不同对象之间的相似性.

例如, 文献[12]提出了一种基于元路径的随机游走策略来生成有意义的节点序列用于网络嵌入. 首先用一组融合函数对学习到的节点嵌入进行变换, 然后将其集成到扩展矩阵分解模型中. 将扩展后的 MF 模型与

融合函数联合优化,用于分级预测任务.考虑到元路径仍不足以反映物品间的交互关系,文献[13]将元图的概念引入基于HIN的推荐.提出了一种基于套索的正则化FM算法,通过自动学习观察到的评级,有效地选择出有用的元图特征.文献[14,15]针对研究学习元路径的表示方式,计算节点 u 和 v 之间存在元路径 M 的概率.第1阶段基于随机游走和负采样生成符合目标关系的数据,用于表示学习.第2阶段用一个神经网络模型,通过最大化预测节点之间关系的可能性,同时学习节点和关系的表示向量.

本文提出的异构信息网络推荐算法与上述研究的不同之处在于:

(1) 现有的GCNs方法,仍未有效利用异构信息网络中复杂的节点和关联信息,其作用对象仍局限于在单一关联类型的同构网络,对于具有异构信息网络特征的真实数据,往往采用简化网络结构的方式来完成训练,损失了大量用户与用户、用户与项目之间的多种交互信息.相比之下,本文的模型关注的就是如何尽可能地利用这些信息,在节点多视角嵌入融合的过程中保留带有高阶行为间交互关系的网络信息,并使用合理的机制为不同的关系分配不同的权重,弥补了这类方法的不足.

(2) 基于图注意力机制的表示模型将图注意力机制引入了节点嵌入模型,实现了对异构信息网络的嵌入.但是现有的方法缺乏对节点之间不同类型的关联的表达,忽视了在不同视角下生成的嵌入也应该具有丰富潜在信息的关联.本文模型吸收了图注意力机制的思想,在嵌入模型的第2阶段使用这一机制对各个视角的嵌入进行关联,连接了不同关系下的邻居节点,提高了模型的效果.

(3) 基于元信息的方法依赖于对元路径或元结构的定义,不同的数据集适合不同的定义,其定义的好坏直接决定了模型效果,所以元信息的方法依赖于人工选择和研究者使用元信息的经验.相比而言,本文模型只设置网络深度,不必固定元结构,在推导式学习时更合适.

本文结合上述3种方案的优点,设计了用于异构信息网络基于多视角嵌入融合的推荐方法,将图注意力机制引入线性卷积网络中,对节点及其关联关系进行嵌入融合,基于协同过滤来设计便于更新的评分模型.

2 MVEF 模型

为了充分利用异构信息网络的多种潜在特征,以提高推荐的准确性,本文提出了一种基于多视角嵌入融合的推荐模型——MVEF.本节首先给出问题定义,然后介绍模型的基本思想.

2.1 问题定义

本文主要解决的是在电商背景下的商品推荐问题,使用电商场景下产生的网络信息,挖掘用户的消费偏好,为用户提供更准确的商品推荐.该模型要解决的问题是如何在已知用户与部分商品之间的偏好关系的前提下,在其他商品之中为用户推荐其更偏好的商品.为此,首先需要对用户及商品之间偏好关系的网络进行定义.现实世界的对象及其交互通常是多种类型的,需要使用一种通用的网络结构来描述现实世界的信息,这种具有多重节点类型和多重边类型的网络,本文称之为异构信息网络.表1总结了本文常用的符号.

表1 本文所使用的符号

名称	描述
$\mathcal{G}, \mathcal{V}, \mathcal{E}$	图, 节点和边的集合
\mathcal{A}, \mathcal{R}	边的类型取值集合
v, e	一个节点, 一条边
n	节点的总数
K	卷积的最大层数
N	所有邻居节点的集合

定义1(异构信息网络). 异构信息网络为具有多种节点类型和边类型的网络图 $G=(\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{R})$, 其中, \mathcal{V} 为 n 个节点的集合,其取值集合为 \mathcal{A} , \mathcal{E} 为节点之间的边的集合,每条边 $e_{ij}=(v_i, v_j)$ 对应了节点 i 和节点 j 之间的关系类型,边的取值集合为 \mathcal{R} ,在本文中, $\mathcal{R}=(r_u, r_i)$,即每个节点 v 具有两种类型 u 和 i .

定义2(高阶同关系邻居节点集). 已知在图 G 中,有目标节点 v ,所有与节点 v 以关系 r 相关联的节点组成

的集合 $\mathcal{V}_r^{(1)}$ 为 1 阶同关系邻居节点集, 则有: 当 $k \geq 2$ 时, 与节点集 $\mathcal{V}_r^{(k-1)}$ 中的任一节点 $v_r^{(k-1)}$ 仍以同一关系 r 相关联、但不属于节点集 $\mathcal{V}_r^{(k-2)}$ 的节点 $v_r^{(k)}$ 组成了集合 $\mathcal{V}_r^{(k)}$, 将这样的 k 阶节点集合称为高阶同关系邻居节点集.

定义 3(同质关联视角下的嵌入融合). 已知图 $G=(\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{R})$ 、关系类型 $r \in \mathcal{R}$ 、传播深度 K , 已知目标节点 v , 节点 v 从 0 到 K 阶的所有高阶同关系邻居节点集 $\mathcal{V}_r^{(k)}$ 中的节点所组成的集合称为 K 阶邻域节点集 $N_{v,r}$. 有嵌入过程 $\phi()$, 则同质关联视角 r 下的嵌入融合为 $v_r = \phi(N_{v,r})$.

定义 4(异质关联视角下的嵌入融合). 已知图 $G=(\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{R})$, 有一组同质关联视角下的节点嵌入所组成的集合 $\hat{V} = \{v_r | r \in \mathcal{R}\}$. 有嵌入过程 $\phi()$, 则异构关联视角下的嵌入融合为 $\bar{v} = \phi(\hat{V})$.

2.2 模型基本思想

MVEF 模型的基本思想如图 2 所示. 本文融合节点在图 G 中的信息, 形成节点在潜在空间的表示 \bar{v}_u, \bar{v}_i , 通过标签数据来优化模型. 更新得到的节点嵌入 \bar{v}_u, \bar{v}_i 就能够表现节点间的相似关系, 从而可以为用户推荐更相关的物品. 本文模型主要使用 GCN 和 GAT 来构建异构信息网络中节点的嵌入, 在卷积传播的过程中, 本文选择只对节点的邻居节点进行平滑, 而不是对整个图的节点进行平滑, 这也符合近年来图卷积推荐算法的研究趋势^[16].

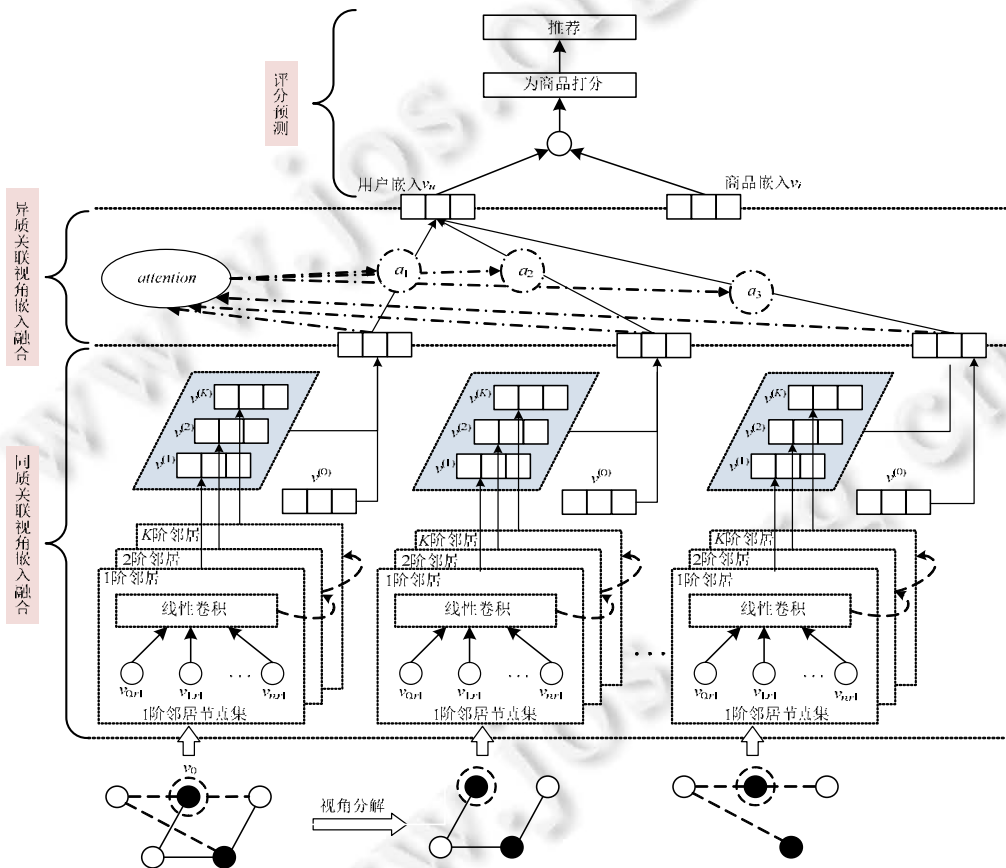


图 2 MVEF 模型的基本思想

本文提出的多视角嵌入融合推荐模型主要分为 3 个阶段, 包括在同质关联视角下的嵌入融合、在异质关联视角下的嵌入融合以及基于评分预测进行推荐. 首先将目标节点的邻域节点集合按不同的关系视角进行划分, 然后进行两阶段的嵌入融合, 第 1 阶段是在同质关联视角下进行局部嵌入融合, 第 2 阶段是在异质关联视角下进行全局嵌入融合. 近期的许多工作解决了同构信息网络中节点如何嵌入的问题^[8], 而在异构信息网络

中,最主要的挑战是节点在对其高阶邻居节点的卷积过程中,要如何有效融合不同属性的信息以及不同关联关系的信息,这是以前的推荐方法中没有解决的问题.而 MVEF 为节点嵌入进行邻域传播的过程划分了不同的视角,每个视角对应一种关联类型,其有效的边使用加粗表示,忽略的边使用虚线表示.

在第 1 阶段中, MVEF 模型使用消息传递结构,通过遮罩层限制每次卷积时只聚合节点的邻居节点,而不是整个图的信息,经过多层卷积后生成了在多个视角下的特征矩阵,这些矩阵融合了在同质关联视角下的节点之间的相似特征信息,但是这些特征之间是相互独立的,使用同质层间池化将它们融合起来. MVEF 模型的第 2 阶段融合了节点在各个同质关联视角下生成的局部嵌入,本文首先通过忽略关系的类型来生成基础的嵌入,然后以基础嵌入为基底,使用注意力机制来连接异质关联视角下的嵌入,从而达到为不同的视角分配权重的目的.

3 多视角嵌入融合策略

本节详细介绍 MVEF 模型是如何从多个视角对目标节点的上下文信息进行嵌入融合的.如图 3 所示, MVEF 模型先按不同关系类型划分不同视角,在各个视角下单独进行局部嵌入融合,再整合全局视角进行全局嵌入融合.

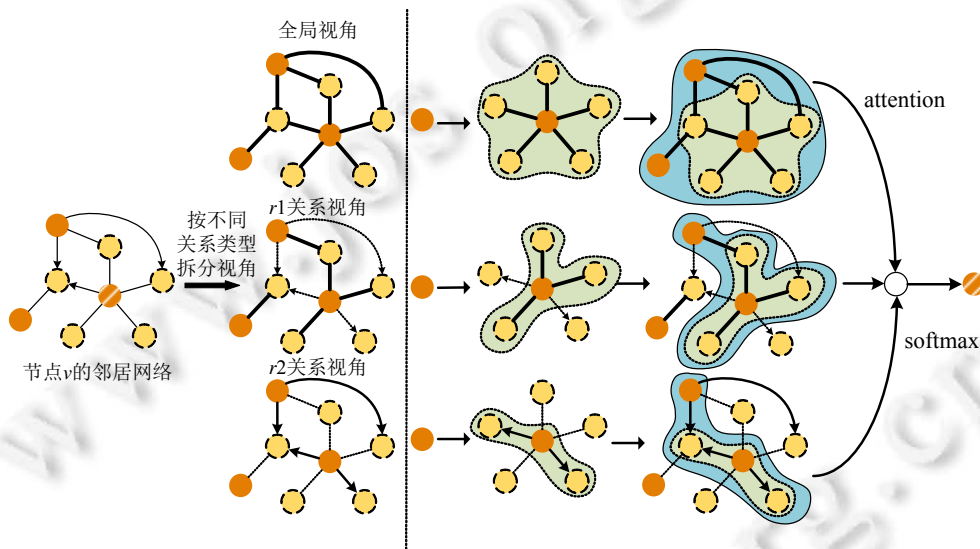


图3 邻域节点信息在多视角下的传播和聚合过程

3.1 同质关联视角下基于GCN的嵌入融合

同质关联视角下嵌入融合的目标是实现节点嵌入在同一关系类型下的局部融合,这一过程由线性卷积和同质层间池化所构成.

3.1.1 线性卷积结构

首先讨论图卷积中的非线性结构在同质关联视角的嵌入融合中是否有必要.如果采用早期提出的图卷积聚合方法^[5],会导致局部和全局两阶段的模型在多级邻域的传播下难以训练,因此需要在保证模型表达能力的同时,尽量使模型复杂度不会随模型规模的线性叠加而呈幂级增加.本文借鉴简单图卷积网络的思想^[17],引入线性的卷积结构聚合邻居节点.

首先是每一层节点只有一个邻居的简单情况(如图 4(a)所示).

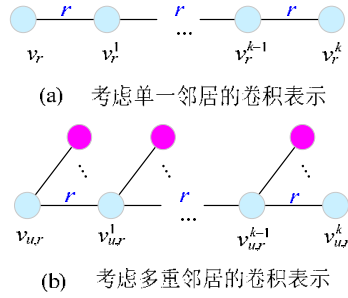


图 4 同质关联视角下基于 GCN 的嵌入融合策略

$$\mathbf{v}_r^{(k+1)} = S\mathbf{v}_r^{(k)}\mathbf{W}_r^k \tag{1}$$

其中, r 为关系类型集合 R 中某一类型, $\mathbf{v}_r^{(k+1)}$ 为节点 v 的 $k+1$ 阶邻居节点的卷积表示, S 为拉普拉斯变换. 对这一卷积过程展开 $k+1$ 次得到:

$$\begin{aligned} \mathbf{v}_r^{(k+1)} &= S\mathbf{v}_r^{(k)}\mathbf{W}_r^k \\ &= SS\mathbf{v}_r^{(k-1)}\mathbf{W}_r^{k-1} \\ &= S\dots S\mathbf{v}_r^{(0)}\mathbf{W}_r^{k0}\dots\mathbf{W}_r^k \\ &= S\dots S\mathbf{v}^{(0)}\mathbf{Y}^k \end{aligned} \tag{2}$$

而经典的图卷积模型每层之间采用非线性结构连接, 如 $\mathbf{v}_r^{(k+1)} = \sigma(\mathbf{W}\mathbf{v}_r^{(k)})$. 若参数矩阵 \mathbf{W} 的复杂度为 $O(|V|)$, 则非线性结构的参数复杂度为 $O(|V|^2)$, 而使用线性的卷积结构, 如式(2)中参数矩阵以点乘运算连接, 于是可以简化为矩阵 $\mathbf{Y}^k: \mathbf{Y}^k = \mathbf{W}_r^0\dots\mathbf{W}_r^k$. 其复杂度从 $O(|V|^k)$ 降低为 $O(|V|)$, 模型复杂度缩小了 $\frac{1}{|V|^k}$. 可见, 去

除了图卷积聚合函数中的非线性结构, 使得需要训练的参数复杂度由 $O(k)$ 降低为 $O(1)$, 大大提高了模型训练的可行性. 基于最近的研究表明, 对于一般的推荐任务, 在节点信息传播阶段去除非线性结构, 不会明显影响推荐的准确度, 且有利于解决过平滑问题^[7], 提高推荐的准确度.

接下来, 在邻居节点个数未知的一般情况下, 设计自连接的节点传播结构. 在这个传播结构中, 首先计算提取同一关系 r 下的邻居节点相应的 k 阶表示 $\mathbf{v}_{i,r}^{(k)}$, 分别乘以该邻居节点的度的倒数, 与当前节点的 k 阶表示相加, 再乘以当前节点的度的倒数, 最后经过一个线性参数矩阵, 就形成了一个线性卷积层, 以用户节点 u 为例(如图 4(b)所示):

$$\mathbf{v}_{u,r}^{(k+1)} = \left[\frac{1}{\sqrt{|N_u|}}\mathbf{v}_{u,r}^{(k)} + \sum_{i \in N_u} \frac{1}{\sqrt{|N_u|}\sqrt{|N_i|}}\mathbf{v}_{i,r}^{(k)} \right] \mathbf{W}_r^k \tag{3}$$

其中, $|N_u|$ 表示节点 u 邻居节点的度, 节点的度代表物品的受欢迎程度, 该物品的受欢迎程度越低, 越能反映这个用户的独特偏好, 这是协同过滤思想的体现, 而在卷积过程中使用自连接结构, 能够更好地保留原点的信息.

3.1.2 同质层间池化

同质层间池化是同质关联视角下嵌入融合的关键步骤, 在实际训练中, 如果直接使用节点的高阶聚合表示 $\mathbf{v}_{u,r} = \mathbf{v}_{u,r}^{(k)}$, 会导致过平滑现象, 严重影响模型的训练效果. 具体来说, 在调整节点传播深度 K 的过程中, 当 K 设置超过某个并不大的阈值时, 节点的表示效果反而会下降. 实际训练过程中, 不同数据集稀疏度不同, 阈值也有所不同, 但一般不大于 2, 导致模型只能聚合周围 2 层的节点, 只包括 1 层商品节点和 1 层用户节点, 信息表达受到严重阻碍. 过平滑问题的出现, 是因为节点的卷积传播过程本质上就是一次平滑运算, 在特殊情况下只是把邻居节点的嵌入累加了起来, 然后取了平均值. 经过多层传播以后, 与原始节点最相近的节点信息反而被多次平滑了, 节点的特征随之丢失. 为此, 本文提出一系列局部嵌入融合策略, 见表 2.

表 2 局部嵌入融合策略

每阶邻居数	单一邻居	多重邻居
多阶邻居是否独立	单一串联策略	多重串联策略
是	单一平均池化策略	多重平均池化策略
否		

为解决这一问题, 很容易联想到将节点每一阶的嵌入串联起来合并为一个嵌入.

$$\mathbf{v}_{u,r} = \mathbf{v}_{u,r}^{(k)} \parallel \mathbf{v}_{u,r}^{(k-1)} \parallel \dots \parallel \mathbf{v}_{u,r}^{(0)} \quad (4)$$

这一方法在最近的研究中也有提及^[7], 但是针对异构信息网络具有丰富潜在信息的特点, 需要设计一种结构, 在尽量不增加模型潜在空间的情况下, 提高模型表达的丰富程度. 这里, 我们提出另一种平均池化结构来聚合各阶下的高阶嵌入.

$$\mathbf{v}_{u,r} = \sum_{k=0}^K a_k \mathbf{v}_{u,r}^{(k)} \quad (5)$$

其中, $\mathbf{v}_{u,r}$ 为节点 u 在关系 r 下的聚合嵌入表示. 当 r 为 0 时连接周围所有节点作为 *base embedding*. K 为超参数, 表示邻域卷积的最大层数, 参数 a_k 统一设置为常数 $C = \frac{1}{K+2}$ 效果较好.

对比上述两种传播优化方案, 假如本文先假设图中只有一种关联类型, 亦即只考虑单一视角下的嵌入融合, 使用点积运算计算两个节点 u 和 i 之间的邻近度 $\tilde{S} = \mathbf{v}_u \cdot \mathbf{v}_i$, 很容易推导得到.

$$\tilde{S}_1 = (\mathbf{v}_u^{(K)} \parallel \dots \parallel \mathbf{v}_u^{(0)}) \cdot (\mathbf{v}_i^{(K)} \parallel \dots \parallel \mathbf{v}_i^{(0)}) = (\mathbf{v}_u^{(K)} \cdot \mathbf{v}_i^{(K)} \parallel \dots \parallel \mathbf{v}_u^{(0)} \cdot \mathbf{v}_i^{(0)}) \quad (6)$$

$$\tilde{S}_2 = C \sum_{k=0}^K \mathbf{v}_u^{(k)} \cdot C \sum_{k=0}^K \mathbf{v}_i^{(k)} = C^2 (\mathbf{v}_u^{(K)} \mathbf{v}_i^{(K)} + \dots + \mathbf{v}_u^{(K)} \mathbf{v}_i^{(0)} + \dots + \mathbf{v}_u^{(0)} \mathbf{v}_i^{(K)} + \mathbf{v}_u^{(0)} \mathbf{v}_i^{(0)}) \quad (7)$$

其中, \tilde{S}_1 是采用串联方法计算的邻近度, \tilde{S}_2 是采用平均池化方法计算的邻近度. 对比 \tilde{S}_1 和 \tilde{S}_2 , 不难发现, 前者有 K 个交叉项, 而后者有 K^2 个交叉项, 使得两种节点每一级的邻居节点都可以得到交互, 显然, 后者的信息表达效果更加丰富. 而从复杂度上分析, 看似增加了传播嵌入的时间, 但其实并没有影响到参数矩阵的大小, 也就保证了传播更新的效率. 因此, 本文采用平均池化法而不是张量串联法, 来保留每一级邻域内的节点信息, 这样在后续的评价估计时才能具有更丰富的表示效果.

综上, 按不同的关系类型为目标节点的邻居分类, 基于线性传播结构, 使用节点度的倒数作为权重的自连接聚合, 结合平均池化方法连接每一层的高维特征, 得到了目标节点在同质关联视角下的嵌入表示.

算法 1. 同质关联视角下的嵌入融合算法 *genHomE*.

输入: 目标节点 $Target_v$, 传播层数 K , 目标节点 $Target_v$ 在关系 r 下的邻居节点集 N_v ;

输出: 同质关联视角 r 下的嵌入 \mathbf{v}_r , 参数矩阵 \mathbf{Y}_r .

1. $E[0] = InitialNode(Target_v)$
2. For $k=1$ to K
3. $N_k_num = countNeighbourNum(Target_v)$
4. For $j=1$ to N_k_num
5. $N_j_num = countNeighbourNum(N_v[k][j])$
6. $hidden[k] = aggreg(hidden, N_v[k])$
7. End For
8. $E[k] = aggreg(hidden[K], N_v[k], E[k-1])$
9. End For
10. $\mathbf{v}_r = meanPool(E[0] \dots E[K])$

Output: \mathbf{v}_r .

3.1.3 算法描述

基于同质关联视角下的嵌入融合模型, 本文提出了同质关联视角下的嵌入融合算法(如算法 1), 主要步骤如下.

步骤 1. 初始化节点嵌入 $E[0]$. 将目标节点及其邻居节点嵌入初始化, 将目标节点的初始嵌入作为 0 阶局部嵌入.

步骤 2. 对 K 阶邻居节点进行局部嵌入. 遍历目标节点各阶邻居, 计算每一阶中所有邻居节点的度, 从局部融合邻居节点的嵌入, 得到局部嵌入 $\{E[0], \dots, E[K]\}$.

步骤 3. 对 K 阶邻居节点的局部嵌入进行融合. 使用平均池化连接各阶的局部嵌入, 构成同质关联视角 r 下融合的局部嵌入 v_r .

3.2 异质关联视角下基于注意力的嵌入融合

同质关联视角下的嵌入融合过程仅考虑了异构信息网络中单一的关联类型, 在局部融合下得到节点的嵌入. 异质关联视角下的嵌入融合是将不同关联类型下的节点嵌入进一步融合, 即实现节点嵌入的全局融合. 为此, 本文提出一系列全局嵌入融合策略(见表 3).

表 3 全局嵌入融合策略

	多视角是否独立	是	否
是否考虑用户行为	是	-	图注意力策略
	否	串联策略	平均池化策略

3.2.1 基于共享注意力的嵌入融合

经过同质关联视角下的嵌入融合, 得到了节点 v 在 n 种不同关系类型下的嵌入集合 $v_set = \{v_{u,r_1}, \dots, v_{i,r_m}\}$, 其中每个嵌入反映了在单一关系类型下当前节点及其上下文的信息. 接下来要将多种关联视角下的嵌入合并为一个全局嵌入, 使这个全局嵌入能够反映节点多种行为之间的关联特征.

本文可以选择串联、池化和注意力机制等多种方式对嵌入进行联合(如图 5 所示). 以用户节点 v_u 为例, 串联和池化的嵌入聚合过程为

$$\bar{v}_u^n = v_{u,1} \parallel v_{u,2} \parallel \dots \parallel v_{u,n} \tag{8}$$

$$\bar{v}_u^n = \sum_{r \in R} \frac{1}{n+1} v_{u,r} \tag{9}$$

其中, \bar{v}_u^n 为串联的聚合方式, 这种方式中多种视角下的嵌入的确整合了起来, 但实际上多种视角之间独立表示, 互不影响, 异质关联视角的嵌入间并没有被关联起来.

而 \bar{v}_u^n 为平均池化的聚合方式, n 为关系的数量. 这种方式下异质视角的嵌入间的确被关联了起来, 但是经过平均池化被分配的权重是相同的, 而节点通过不同关系进行连接, 其代表的信息却是不同的. 从现实场景出发, 对于用户的喜好的表示, 用户所购买的商品应该比其所浏览的商品贡献更大的信息比重. 在同质关联视角下使用平均池化, 是因为在每一阶的嵌入过程中已经携带了节点的度的信息, 可以反映节点与节点之间重要性的区别, 而在异质视角下就需要采取类似的机制来分配权重.

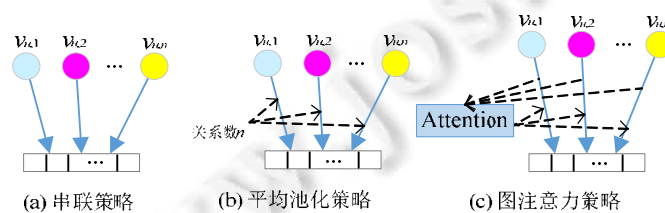


图 5 异质关联视角下的嵌入融合策略

综上所述, 本文采用使用图注意力机制为不同视角下的嵌入分配权重, 将同质关联视角下的嵌入进行联合计算, 得到该节点最终的表示:

$$\begin{aligned}
\bar{\mathbf{v}}_u &= \sigma \left(\sum_{j=0}^n \alpha_{u,j} \mathbf{v}_{u,r_j} \right) \\
&= \sigma \left(\bar{a} \sum_{j=0}^n \alpha_{u,j} (\mathbf{v}_{u,j}^{(0)} + \dots + \mathbf{v}_{u,j}^{(K)}) \right) \\
&= \sigma \left(\bar{a} \alpha_{u,0} (\mathbf{v}_{u,0}^{(0)} + \dots + \mathbf{v}_{u,0}^{(K)}) + \dots + \bar{a} \alpha_{u,n} (\mathbf{v}_{u,n}^{(0)} + \dots + \mathbf{v}_{u,n}^{(K)}) \right)
\end{aligned} \tag{10}$$

其中, $\frac{1}{K+1}$ 记为 \bar{a} , $\sigma(\cdot)$ 为激活函数, 选择 LeakyReLU 函数. 一个节点经过展开会得到 $n \times K$ 项, 如果两个节点再经过点积计算就会得到 $(n \times K)^2$ 项, 因此本文将每个节点同一关联类型下的注意力系数共享, 在每种关系 l 的视角下分别计算注意力系数:

$$\mathbf{e}_{u,j}^l = \text{LeakyReLU}(\bar{\mathbf{w}}^l \|\mathbf{v}_u^l\| \mathbf{v}_j^l) \tag{11}$$

$$\alpha_{u,i}^l = \frac{\exp(\mathbf{e}_{u,i}^l)}{\sum_{j \in N_u} \exp(\mathbf{e}_{u,j}^l)} \tag{12}$$

$$\alpha_{u,i} = (\alpha_{u,i}^0 \|\alpha_{u,i}^1\| \dots \|\alpha_{u,i}^l\|) \tag{13}$$

其中, \mathbf{v}_u^l 即 *base embedding*. 使用共享的注意力后, 注意力系数的复杂度由 $O((n \times K)^2)$ 降低到 $O(n \times K)$, 缩小了模型规模, 提高了计算的可行性.

3.2.2 算法描述

基于上述思想, 提出了异质关联视角下的嵌入融合算法, 其过程如算法 2 所示. 算法 2 主要步骤如下.

步骤 1. 全局关联视角嵌入融合. 不区分关联类型, 融合目标节点邻居的嵌入, 生成的全局关联视角嵌入融合 \mathbf{v}_0 用作注意力机制连接的基点.

步骤 2. 基于异质关联视角计算注意力结构. 将各个同质关联视角下生成的局部嵌入 $\dot{\mathbf{V}} = \{\mathbf{v}_r \mid r \in \mathbf{R}\}$ 以全局关联视角嵌入 \mathbf{v}_0 作为基底, 计算 $\alpha = \{\alpha_r \mid r \in \mathbf{R}\}$, 并对同质关联视角下的局部嵌入 $\dot{\mathbf{V}}$ 与注意力矩阵 α 进行矩阵乘法, 经过一个非线性变换后得到目标节点在异质关联视角下的全局嵌入 $\bar{\mathbf{v}}$.

算法 2. 异质关联视角下的嵌入融合算法 *genHetE*.

输入: 目标节点 *Target_v*, 目标节点 *Target_v* 在不同视角下生成的嵌入集 $\dot{\mathbf{V}} = \{\mathbf{v}_r \mid r \in \mathbf{R}\}$;

输出: 异质关联视角下目标节点的嵌入 $\bar{\mathbf{v}}$, 注意力系数 α .

1. $\mathbf{v}_0 = \text{genBase}(\text{Target}_v)$ //生成全局视角嵌入
2. For $r=1$ to R
3. $\text{value}_r = \text{LeakyReLU}(w, \mathbf{v}_0, \mathbf{v}_r)$
4. $\alpha_r = \text{softmax}(\text{value}_r, \mathbf{v}_0)$ //经过 softmax 计算注意力系数
5. $\text{hidden} = \text{hidden} + \text{Mul}(\alpha_r, \mathbf{v}_r)$
6. End For
7. $\bar{\mathbf{v}} = \text{aggre}(\alpha, \text{hidden})$

Output: $\bar{\mathbf{v}}$

4 基于多视角嵌入融合的推荐算法

本节介绍该推荐方法是如何使用多视角嵌入融合模型来进行推荐预测的.

4.1 评分计算

对于异构信息网络的常见推荐场景, 希望能够从用户的多种交互记录中挖掘出用户之间相似的兴趣, 并确定每个用户自身的偏好. 因此本文采用相似度评分方法, 将两个节点 u 和 i 之间的相似度作为二者的评分依据. 假设两个节点之间存在实际关联或潜在关联, 那么它们的嵌入之间越相似, 它们也就越相关, 本文将 u 、 i 两个节点之间的关联强度称为 $\phi(u, i)$, 因此,

$$\varphi(u, i) = \text{Sim}(u, i) \quad (14)$$

本文用多视角嵌入融合模型得到的节点嵌入 \bar{v}_u 、 \bar{v}_i 表示计算节点之间的相似度 $\text{Sim}(u, i)$, 节点之间的相关评分就可以近似看作节点之间的关联强度:

$$\widehat{r}_{u,i} = \langle \bar{v}_u, \bar{v}_i \rangle = \sigma \left(\bar{a} \sum_{j=0}^n \alpha_{u,j} \left(\mathbf{v}_{u,j}^{(0)} + \dots + \mathbf{v}_{u,j}^{(K)} \right) \right) \cdot \sigma \left(\bar{a} \sum_{j=0}^n \alpha_{i,j} \left(\mathbf{v}_{i,j}^{(0)} + \dots + \mathbf{v}_{i,j}^{(K)} \right) \right) \quad (15)$$

$\langle \cdot \rangle$ 为点积操作, 由于 \bar{v}_u 和 \bar{v}_i 内部采用平均池化连接, 因此每一个嵌入分量之间都会产生交互的权值矩阵, 从而构建出丰富的特征空间.

4.2 模型训练

接下来讨论如何对提出的嵌入推荐模型进行训练. 本文基于随机游走算法思想在节点序列上采样节点及其上下文获取嵌入, 采样过程完全随机. 本文的数据集是用户-商品二部图 G , 给定图 G 中的视角 r , 图 $G_r = (\mathcal{V}, \mathcal{E}_r, \mathcal{A})$, 随机游走过程可以描述为 $\mathcal{T}: \mathcal{V}_1 \rightarrow \dots \rightarrow \mathcal{V}_t \rightarrow \dots \rightarrow \dots \rightarrow \mathcal{V}_K$, 其中, K 为最大的游走长度, 即网络的最大深度, 每一步 t 的转移概率定义为

$$p(v_j | v_i, \mathcal{T}) = \begin{cases} \frac{1}{|\mathcal{N}_{i,r} \cap \mathcal{V}_{t+1}|}, & (v_i, v_j) \in \mathcal{E}_r, v_j \in \mathcal{V}_{t+1} \\ 0, & (v_i, v_j) \in \mathcal{E}_r, v_j \notin \mathcal{V}_{t+1} \\ 0, & (v_i, v_j) \notin \mathcal{E}_r \end{cases} \quad (16)$$

其中, $v_i \in \mathcal{V}_t$, $\mathcal{N}_{i,r}$ 表示节点 v_i 在关系 r 上的邻域节点集合. 经过随机游走, 得到节点的上下文集合 C , 如果给定节点集合 \mathcal{V}_t 及其所有邻域节点, 一个通用的多任务目标函数结构 $\mathcal{L}(\Theta)$ 为

$$\mathcal{L}(\Theta) = \sum_{k=1}^K \lambda_k \tilde{\mathcal{L}}_k(\Theta_1) + \mu \|\Theta_2\|^2 \quad (17)$$

其中, $\Theta = \{\Theta_1, \Theta_2\}$, Θ_1 为嵌入及推荐结构中的参数矩阵, Θ_2 为正则化系数, 防止过拟合, λ_k 为参数矩阵. 得到本文 MVEF 模型的目标函数为

$$\min_{\Theta} \mathcal{L}(\Theta) = - \sum_{u=1}^N \sum_{i \in \mathcal{N}_u} \sum_{j \in \mathcal{N}_u} \ln \sigma(\hat{r}_{u,i} - \hat{r}_{u,j}) + \mu \|\Theta_2\|^2 \quad (18)$$

对 N 个目标节点进行采样, 生成样本总体, 在实验时根据预设的样本比例, 如 20%、100% 等对采样的正负例进行交叉计算, 鼓励观察到的样本高于未观察到的样本, 使用 Adam 优化进行小批量的优化. 由于模型中非线性结构的比例较低, 且引入了 $L2$ 惩罚项, 在 $k \leq 3$ 的传播深度以下并未出现过拟合问题.

4.3 算法描述

本节提出了一种基于多视角嵌入融合的推荐算法, 其过程如算法 3 所示.

算法 3 主要步骤如下.

步骤 1. 数据初始化. 在图中随机游走, 采样生成目标节点集合 \mathcal{N}_0 和邻居节点集合 \mathcal{N}_v .

步骤 2. 对目标节点进行多视角嵌入融合. 遍历目标节点集合, 对目标节点的多种关系进行迭代, 调用 *genHomE* 算法计算各个同质关联视角下的局部嵌入, 调用 *genHetE* 算法计算异质关联视角下的全局嵌入.

步骤 3. 迭代更新模型. 计算任意用户和商品节点间的推荐评分并计算损失, 迭代更新参数空间, 直到小于预期损失.

步骤 4. 计算推荐列表. 根据生成的嵌入融合模型, 为用户集合计算推荐评分最高的商品集合 *RecList*.

算法 3. 基于多视角嵌入融合的推荐算法.

输入: 目标节点集 \mathcal{N}_0 , 传播层数 K , 学习率 a ;

输出: 目标节点集的推荐列表 *RecList*.

1. $\mathcal{N}_v = \text{deepwalk}(\mathcal{N}_0)$
2. While (*Loss* < *minLoss*) // 迭代更新

```

3.   For  $i=1$  to  $N$ 
4.     For  $r=1$  to  $R$ 
5.        $homE[i][r]=genHomE(\mathcal{N}_0[i], r)$ 
6.     End For
7.      $genHetE(homE[i])$ 
8.   End For
9.    $Loss=countLoss(\{\bar{v}_u, \bar{v}_i \mid u, i \in \mathcal{N}_0\})$ 
10.   $\{Y, w, \alpha\}=updateWeights(Loss, a)$ 
11. End While //迭代完毕
12. For  $i=1$  to  $N$ 
13.   $RecList[i]=maxRateItem(\mathcal{N}_0[i])$ 
14. End For
Output:  $RecList$ 

```

5 实验

本节介绍实验所用的数据集和实验设置、评价指标以及相关的对比实验。

5.1 数据集

本实验采用两个真实的电子商务数据集: Amazon^[18]和 Taobao User Behavior 数据集, 这两个数据集中均可提取出 3 种不同的用户行为, 包括“收藏”“加入购物车”和“购买”。本文对这两个数据集进行预处理, 去除掉不必要的标签, 过滤掉交互次数过少且不具有代表性的节点(至少存在 5 次交互记录), 并将剩下的数据划分为训练集、验证集和测试集, 处理后的数据集统计情况见表 4。

表 4 评估数据集细节统计

Dataset	User	Item	Behavior1	Behavior2	Behavior3
Amazon	31 524	12 895	45 855	72 347	52 653
Taobao	48 732	39 426	1 548 011	192 156	258 652

5.2 评价指标

本文采用两个经典的指标 HR 和 NDCG 对推荐的性能进行度量。

(1) 命中率(HR), 这是一个基于召回的指标, HR 通过计算所测试的目标项目是否在推荐生成的 Top-K 列表中, 来反映系统是否推荐到了用户想要的商品, 从而表现预测的命中率:

$$HR = \frac{1}{N} \sum_{i=1}^N hits(i) \quad (19)$$

其中, N 表示被推荐的用户的总数量, $hits(i)$ 表示第 i 个用户访问的值是否在推荐列表中。

(2) 归一化折损累计增益(NDCG), 这是对排序结果的评价指标, 是对折损累计增益(NDCG)的一种改进方式。由于不同的用户其列表长度各不相同, 所以不同用户的 NDCG 比较并不准确, 因此需要对不同用户的指标进行归一化。使用 NDCG 可以反映所推荐的项目是否在用户更加显眼的位置上, 从而表现预测的精准性。

$$NDCG = \frac{1}{N} \sum_{i=1}^N \frac{1}{\log(p_i + 1)} \quad (20)$$

其中, p_i 表示第 i 个用户的实际访问商品在所生成的推荐列表中的位置。使用 HR 和 NDCG 能够从推荐结果的命中率和排序性角度综合反映推荐的准确性。

5.3 参数设置

本文模型中有几个重要的参数: 学习率、传播的深度 K 、节点嵌入矩阵 V 的维度 D , 以及目标函数中的

正则化参数 L_2 .

我们通过设置不同的学习率、正则化系数、嵌入维度以及传播深度来进行实验. 如图 6(a)和图 6(b)所示, 当学习率为 0.001、 L_2 正则化参数为 0.001 时推荐结果最准确. 如图 6(c)所示, 推荐效果随模型嵌入维度的增大而提高, 当维度从 64 增大为 128 时效果没有明显提升, 但是训练时间和内存占用大大增加, 所以本实验设置各类节点的嵌入维度为 64. 对传播深度进行测试, 在 20%采样水平下, K 从 0 到 5 分别进行实验, 如图 7 所示, 当传播深度从 0 提高到 3 时, 推荐效果逐渐提高; 当传播深度大于 3 时, 准确度提升并不明显甚至有所下降, 并伴随训练时间及模型空间占用的大大增加, 所以本文选择的传播深度为 3 层.

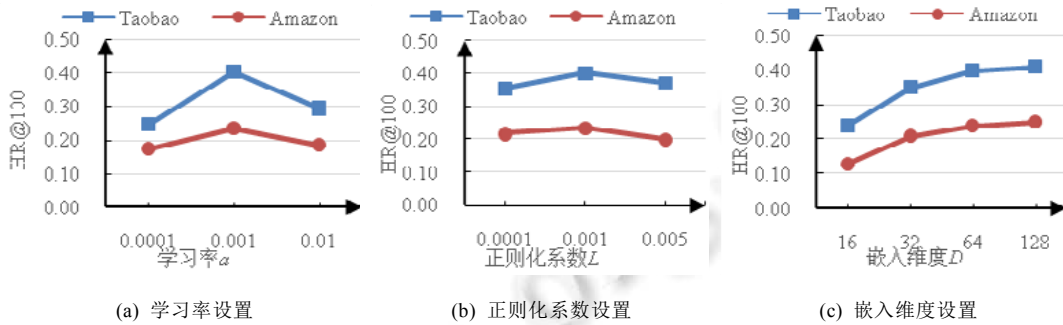


图 6 不同参数设置时的效果对比

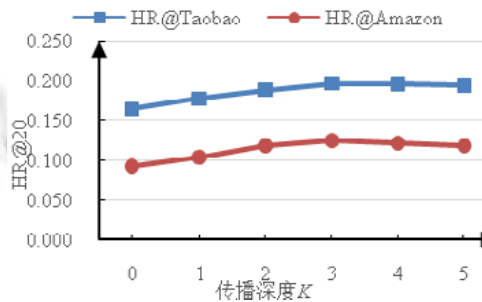


图 7 不同传播深度 K 下的性能比较

传播的深度影响了模型的复杂度, 理论上, 传播的层数越多, 节点的嵌入就聚合了越多的邻居节点信息, 潜层空间的表现力就越强. 但从实践角度来看, 深层节点嵌入容易发生平滑问题, 这一点发生的原因在相关研究中有讨论^[14,19]. 总的来说, 随着层数的增多, 低阶邻居的系数逐渐减小, 而高阶邻居的系数逐渐增大. 注意力机制有助于解决注意力问题, 只是本文的模型在同质视角下的模型建构并没有使用注意力机制, 因此仍然产生了平滑问题. 但是, 由于在本文的模型中选择 $K=3$ 已经有较好的效果, 因此没有扩大模型深度的需要.

5.4 性能评估

为了测试 MVEF 模型的性能, 下面列举几个经典的基准方法和几个近期效果突出的方法, 并将本文提出的方法与这些方法进行对比实验.

- (1) BPR, 基于贝叶斯理论的推荐模型, 这是一种广泛用于项目推荐的学习方法^[20].
- (2) MC-BPR, 该方法将 BPR 模型应用于异构图数据时去除了降采样^[21].
- (3) LightGCN, 这是一种轻量化的图卷积嵌入模型, 去除了图卷积中的非线性结构, 提出了隐含层间聚合的结构^[8].
- (4) GATNE, 该方法将图注意力机制用于异构图数据, 将节点和边的嵌入分别计算生成聚合^[10].

在不同的采样规模下,将本文提出的 MVEF 推荐算法与上述方法进行了对比实验,在 Amazon 数据集和 Taobao 数据集上分别进行训练和测试,采样实验结果见表 5 和表 6.实验结果表明,MVEF 在 Amazon 和 Taobao 数据集上比现有的基准方法具有更好的推荐效果.

表 5 MVEF 与其他方法的性能比较(Amazon 数据集)

模型	HR@20	NDCG@20	HR@40	NDCG@40	HR@60	NDCG@60	HR@80	NDCG@80	HR@100	NDCG@100	
同构	BPR	0.088 2	0.034 1	0.122 1	0.053 4	0.153 4	0.071 1	0.189 1	0.080 3	0.236 4	0.093 8
	LightGCN	0.156 1	0.063 5	0.188 4	0.075 4	0.224 2	0.091 7	0.268 9	0.106 3	0.278 4	0.114 5
异构	MC-BPR	0.134 6	0.053 3	0.174 4	0.067 9	0.239 1	0.081 2	0.283 1	0.113 4	0.346 6	0.132 5
	GATNE	0.167 5	0.077 9	0.210 9	0.089 0	0.275 7	0.101 7	0.329 2	0.131 7	0.367 9	0.145 6
	MVEF	0.196 7	0.077 1	0.248 1	0.104 2	0.317 6	0.126 1	0.356 6	0.142 5	0.401 7	0.160 9

表 6 MVEF 与其他方法的性能比较(Taobao 数据集)

模型	HR@20	NDCG@20	HR@40	NDCG@40	HR@60	NDCG@60	HR@80	NDCG@80	HR@100	NDCG@100	
同构	BPR	0.059 4	0.023 9	0.094 2	0.031 7	0.133 4	0.041 6	0.149 0	0.160 8	0.065 3	
	LightGCN	0.086 2	0.034 6	0.119 2	0.045 3	0.150 7	0.050 2	0.174 7	0.061 3	0.185 3	0.074 1
异构	MC-BPR	0.104 7	0.042 7	0.125 1	0.059 8	0.164 7	0.064 9	0.189 6	0.070 1	0.200 4	0.081 1
	GATNE	0.121 1	0.049 8	0.134 8	0.054 6	0.172 4	0.069 1	0.204 5	0.083 7	0.217 6	0.087 9
	MVEF	0.125 1	0.050 4	0.152 6	0.062 4	0.199 3	0.069 8	0.219 8	0.081 2	0.235 1	0.093 0

为了验证模型结构的合理性,对 MVEF 模型进行不同模型结构的对比实验.在两阶段的嵌入聚合模型中,MVEF 在第 1 阶段采用了简化的图卷积神经网络,在第 2 阶段采用了图注意力网络,假如只考虑图卷积和注意力机制这两种结构,可以得到如下 4 种不同的组合方式.

(1) MVEF-CC,这种方式可以看作经典的 GCN 方法在异构信息网络上的应用,与 MVEF 的区别在于:第 2 阶段采用了使用平均池化的网络进行连接,而没有为不同的关系分配不同的权重.

(2) MVEF-CA,即本文的 MVEF 模型.

(3) MVEF-AC,这种方式近似于 GATNE 模型,但在邻居节点聚合过程中分配权重对实验效果的提升有限,而这种方法忽视了对不同关系视角进行权重分配的重要性.

(4) MVEF-AA,这种方式考虑了两阶段权重的分配,但是两阶段都分配注意力系数,参数空间过于庞大,实验的计算设备难以承受,训练时间过长,模型难以收敛,本文方法可以看作这种方法的优化版本.但是,由于第 4 种结构难以训练,不具有实际使用意义,所以选择前 3 种结构在 Amazon 和 Taobao 数据集上进行训练效果对比.

实验结果如图 8(a)和图 8(b)所示,横坐标表示采样水平,柱的高度对应 HR 值.从图中可以看出,对于两种数据集,MVEF-CA 方法效果都好于 MVEF-AC 和 MVEF-CC.

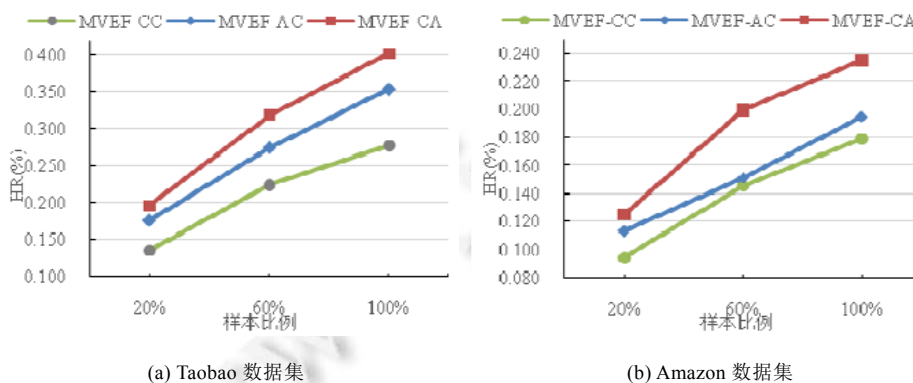


图 8 不同模型架构训练效果对比

6 总 结

针对现有推荐方法研究的不足,我们对异构信息网络的视角嵌入融合推荐方法进行了研究.本文充分分析了异构信息网络的特征,提出了一种基于多视角嵌入融合模型,分两个阶段从同质关联视角和异质关联视角对异构信息网络的深层潜在特征进行挖掘并加以融合,在同质关联视角下使用轻量的图卷积对节点邻域的高阶上下文信息进行局部的嵌入融合,在异质关联视角下使用注意力机制来分配不同关联类型对节点嵌入的影响权重,从而构建节点嵌入的全局融合.此外,本文提出了一种基于多视角嵌入融合模型的推荐算法,实验结果表明:该模型在两个大型的数据集上取得了推荐效果的提升,实验还验证了本文中所提出的关键技术的可行性和有效性.下一步,我们将考虑如何在更大规模的图结构数据上进行模型的搭建和训练.

References:

- [1] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. In: Proc. of the 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2014. 701–710.
- [2] Grover A, Leskovec J. node2vec: Scalable feature learning for networks. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2016. 855–864.
- [3] Dong Y, Chawla NV, Swami A. Metapath2vec: Scalable representation learning for heterogeneous networks. In: Proc. of the 23rd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2017. 135–144.
- [4] Ying R, He R, Chen K, *et al.* Graph convolutional neural networks for Web-scale recommender systems. In: Proc. of the 24th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. 2018. 974–983.
- [5] Wang X, He X, Wang M, *et al.* Neural graph collaborative filtering. In: Proc. of the 42nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 2019. 165–174.
- [6] Berg R, Kipf TN, Welling M. Graph convolutional matrix completion. arXiv preprint arXiv: 1706.02263, 2017.
- [7] Chen L, Wu L, Hong R, *et al.* Revisiting graph based collaborative filtering: A linear residual graph convolutional network approach. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2020. <https://arxiv.org/pdf/2001.10167v1.pdf> [doi: 10.1609/aaai.v34i01.5330]
- [8] He X, Deng K, Wang X, *et al.* LightGCN: Simplifying and powering graph convolution network for recommendation. In: Proc. of the 43rd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 2020. 639–648.
- [9] Veličković P, Cucurull G, Casanova A, *et al.* Graph attention networks. arXiv preprint arXiv: 1710.10903, 2017.
- [10] Cen Y, Zou X, Zhang J, *et al.* Representation learning for attributed multiplex heterogeneous network. In: Proc. of the 25th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. 2019. 1358–1368.
- [11] Fu X, Zhang J, Meng Z, *et al.* MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding. In: Proc. of the Web Conf. 2020. 2331–2341.
- [12] Shi C, Hu B, Zhao WX, *et al.* Heterogeneous information network embedding for recommendation. IEEE Trans. on Knowledge and Data Engineering, 2018, 31(2): 357–370.
- [13] Zhao H, Yao Q, Li J, *et al.* Meta-graph based recommendation fusion over heterogeneous information networks. In: Proc. of the 23rd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2017. 635–644.
- [14] Fu T, Lee WC, Lei Z. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In: Proc. of the 2017 ACM on Conf. on Information and Knowledge Management. 2017. 1797–1806.
- [15] Yang C, Xiao Y, Zhang Y, *et al.* Heterogeneous network representation learning: Survey, benchmark, evaluation, and beyond. arXiv preprint arXiv: 2004.00216, 2020.
- [16] Wang X, He X, Wang M, *et al.* Neural graph collaborative filtering. In: Proc. of the 42nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 2019. 165–174.
- [17] Wu F, Souza A, Zhang T, *et al.* Simplifying graph convolutional networks. In: Proc. of the Int'l Conf. on Machine Learning. PMLR. 2019. 6861–6871.

- [18] Ni J, Li J, Mcauley J. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019. 188–197.
- [19] Chen D, Lin Y, Li W, *et al.* Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. arXiv: 1909.03211, 2019. [doi: 10.48550/arXiv.1909.03211]
- [20] Qiu H, Liu Y, Guo G, *et al.* BPRH: Bayesian personalized ranking for heterogeneous implicit feedback. Information Sciences, 2018, 453: 80–98.
- [21] Loni B, Pagano R, Larson M, *et al.* Bayesian personalized ranking with multi-channel user feedback. In: Proc. of the 10th ACM Conf. on Recommender Systems. 2016. 361–364.



石乐昊(1997—), 男, 硕士生, CCF 学生会员, 主要研究领域为机器学习, 推荐系统.



聂铁铮(1980—), 男, 博士, 副教授, CCF 高级会员, 主要研究领域为大数据管理, 数据集成与融合, 区块链.



寇月(1980—), 女, 博士, 副教授, CCF 专业会员, 主要研究领域为社交网络分析与挖掘, 个性化推荐, 数据集成与融合.



李冬(1979—), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为机器学习, 社交网络.



申德荣(1964—), 女, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为分布式数据管理, 大数据集成, 社交网络, 知识图谱.