信息处理和57管理(2020) 102277

内容列表可在ScienceDirect上找到

信息处理和管理

杂志主页：www.elsevier.com/locate/infoproman

MGAT: 用于推荐的多模态图形注意力网络

陶竹林[a], 魏银伟[b], 王翔[c], 贺湘南, 黄祥林[*,a], 达生-卓克

[a] 中国传媒大学媒体融合与传播国家重点实验室，北京，中国
[b] 山东大学，中国
[c] 新加坡国立大学，新加坡
中国科技大学，合肥，中国

我的朋友们，你们知道吗？

关键词。
个性化推荐 图表
门机制 注意机制 微视频

A B S T R A C T

图形神经网络（GNNs）在个性化推荐方面显示出巨大的潜力。其核心是将交互数据重组为一个用户-物品的二方图，并利用用户和物品节点之间的高阶连接来丰富它们的表现。虽然取得了巨大的成功，但大多数现有的工作只考虑了基于ID信息的交互图，而忽略了来自多种模式的项目内容（*例如，*微视频项目的视觉、听觉和文本特征）。在最近提出的MMGCN（Wei等人，2019年）之前，还没有对不同模式的个人兴趣进行细化区分的研究。然而，它只是在平行的交互图上采用了GNN，并平等地对待从所有邻居传播的信息，未能自适应地捕捉用户的偏好。因此，获得的表征可能会保留再冗余，甚至是嘈杂的信息，导致非鲁棒性和次优的性能。在此工作中，我们旨在研究如何在多模态交互图上采用GNNs，以适应性地捕捉用户在不同模态上的偏好，并对一个项目为什么适合用户进行深入分析。为此，我们提出了一个新的多模态图注意网络，简称MGAT，它在模态的颗粒度上分解个人兴趣。在多模态交互图的基础上，MGAT在单个图中进行信息传播，同时利用门控注意力机制来识别不同模态对用户偏好的不同重要性。因此，它能够捕捉隐藏在用户行为中的更复杂的交互模式，并提供更准确的再赞扬。在两个微视频推荐数据集Tiktok和MovieLens上的实证结果表明，MGAT比最先进的基础数据集有很大的改进。

线，如NGCF（Wang, He, et al., 2019）和MMGCN（Wei et al., 2019）。对一个案例的进一步分析说明了MGAT是如何在多模态行动间图上产生周到的信息流的。

1. 简介

个性化推荐已经成为许多面向用户的服务中的关键组成部分，准确及时地过滤用户感兴趣的项目，对于电子商务中的产品（*如*亚马逊和淘宝）、社交网络中的朋友（*如。*

通讯作者。

电子邮件地址： taozhulin@gmail.com（Z. Tao），weiyinwei@hotmail.com（Y. Wei），xiangwang@u.nus.edu（X. Wang），xiangnanhe@gmail.com（X. He），huangxl@cuc.edu.cn（X. Huang），chuats@comp.nus.edu.sg（T.-S. Chua）。

https://doi.org/10.1016/j.ipm.2020.102277

2020年17月收到；2020年415月收到修订版；4月20接受 2020

可在线阅读125月 2020

Facebook和微信），以及内容分享（如Tiktok和Kwai）平台的微视频。因此，预测用户与一个项目的互动（如购买、点击和查看）的可能性是至关重要的。为此，现有的推荐模型（Chen等人，2012；He & Chua，2017；He等人，2017；Koren，Bell，& Volinsky，2009；Rendle，Freudenthaler，Gantner，& Schmidt-Thieme，2009；Wang，He，Wang，Feng，& Chua，2019）主要侧重于利用历史用户与物品的互动来进行预测。这些模型主要遵循一个一般的范式，配备了两个关键部分--（1）表征学习，将每个用户-物品对及其侧面信息转换为适当的表征，以及（2）交互建模，根据表征进行预测。例如，作为早期的工作，矩阵分解（MF）（Koren等人，2009；Rendle等人，2009）只是将用户（或物品）的ID投射到他/她的嵌入中；后来，FISM（Kabbur，Ning，& Karypis，2013）和SLIM（Ning & Karypis，2011）将历史物品的嵌入平均值作为用户的代表；此外，SVD++（Chen等人，2012）将用户的ID嵌入和历史项目聚合在一起，而NAIS（He，He等人，2018）在历史项目上采用关注机制，以实现更好的性能。显然，表示质量是影响推荐模型有效性的一个关键因素。

最近的研究（van den Berg，Kipf，& Welling，2017；Wang，He，Cao，Liu，& Chua，2019；Wang，He，Wang，et al.，2019；Zheng，Lu，Jiang，Zhang，& Yu，2018）表明，采用图神经网络（GNNs）能够增强用户和物品之间高阶关系的表征学习。这些模型将用户和微视频之间的互动组织成双线图，将它们的关系表现为它们的连接性。在双子图中，一阶连接（即直接连接）呈现了用户和项目的预先存在的特征（例如，历史项目被视为用户档案），而高阶连接反映了用户之间的行为相似性，项目之间的受众相似性，以及协作过滤信号。这样的高阶连通性对于丰富用户和项目的表征非常有用。同时，受GNN的信息传播启发，这些推荐者采用同样的思路来完善用户和物品的表征--即首先生成来自每个邻居的信息，然后汇总所有邻居的信息来更新用户和物品节点的嵌入，并递归执行这种传播来考虑高跳邻居。例如，GC-MC（van den Berg等人，2017）、NGCF（Wang，He，Wang等人，2019）和LightGCN（He等人，2020）都受益于这种传播，其中NGCF将协同过滤（CF）信号编码为表征，并实现了最先进的性能，而LightGCN进一步简化了NGCF的神经网络设计，显示出更好的CF效果。此外，基于GNN的推荐器在许多具有挑战性的场景中显示出巨大的潜力，从社交（Fan等，2019；Wu，Sun等，2019；Wu，Zhang等，2019）、基于会话（Song等，2019；Wu等，2019c；Zheng，Gao，He，Li，& Jin，2020）到基于知识图谱（Wang，Zhao，Xie，Li，& Guo。
2019;Wang，He，Cao，et al.，2019）推荐。

虽然取得了巨大的成功，但这些方法不足以为具有多模态内容的项目建立令人满意的表征，例如Netflix的电影和Tiktok的微视频，它们通常涉及视觉、声音和文本内容（Wei等人，2019）。一个关键原因是，基于GNN的表征学习缺乏对模态差异的明确建模，而模态差异在用户与项目的互动中是潜在的，对于在模态的颗粒度上传播个人兴趣至关重要。更具体地说，大多数现有的方法要么通过将多模态内容视为节点特征来构建一个交互图（van den Berg等人，2017年），要么基于单个模态并行分析多个交互图（Wei等人，2019年），而没有对不同模态的用户口味进行分类，这可能对图上的信息传播产生不同的影响。以微视频推荐为例，用户可能更喜欢具有相同BGM（背景音乐）的微视频，以适应场景的情绪，而场景可能在视觉上有所不同；或者，她可能更喜欢某些微视频的BGM，而更关心其他微视频的视觉场景。因此，同质化或统一的多模态渠道不足以识别不同模态的重要性，阻碍了信息传播并导致次优的表现。

在这项工作中，我们旨在研究如何正确地利用GNN，并在多模态交互图上有效地传播信息，同时捕捉到用户对不同模态的偏好。为此，我们提出了一个新的多模态图注意网络，被称为MGAT，它配备了三个设计。(1）多模态交互图的构建，用于捕捉用户对不同模态的细粒度偏好，这与之前的努力（MMGCN Wei et al.(2019)），在用户和物品节点之间建立平行图；（2）在单个图上进行嵌入传播，用于将用户的行为模式编码到用户和物品的表征中，根据用户的历史物品（或其用户组）更新用户（或物品）的表征；（3）跨图的门控注意力聚合，用于识别不同模态的不同重要性，利用其他模态来学习每个邻居的权重，进而指导传播。因此，这样的注意力机制赋予了MGAT在模式的颗粒度上拆分个人兴趣的能力。此外，在反复进行这种信息传播时，我们可以从高阶邻居那里获得信息流，并根据注意力合理地探索用户的兴趣。我们在两个真实世界的数据集Tiktok和MovieLens上进行了广泛的实验，以证明我们MGAT模型的合理性和有效性。

请注意，这项工作的初步版本已经作为会议论文发表在ACM MM（2019Wei等人，2019）。我们将主要的改进措施总结如下。

- 我们将MMGCN的框架增强为MGAT。相对于MMGCN利用原始GCN可能导致重叠信息冲突并采用最高阶表示法进行预测，MGAT采用GNN对邻接节点的信息进行聚合，并将聚合结果与头部实体节点的信息相结合。同时，MGAT将不同阶数的节点表示进行分类，以区分不同阶数对交互预测的不同贡献。

在MGAT中，我们引入了门控注意力机制，以控制和加权每个模式的每一层中的信息流传播。

- 我们补充了除基线之外的所有实验，以证明我们重建的模型和提出的方法的有效性。
chanism.
- 我们重新组织了论文，以强调这个扩展版本的动机。简而言之，主要贡献归纳如

下。

- 我们开发了一种新的方法MGAT，它将注意力机制纳入图神经网络框架，以分解用户对不同模式的偏好。
- 在技术上，该模型引入了门控注意力机制，对多模态交互图中的信息流进行控制和加权，从而促进了对用户行为的理解。
- 我们在两个数据集上进行了广泛的实验，以验证MGAT的合理性和有效性。此外，由于用户的隐私问题，在这项工作中只考虑用户的ID。我们将在接受后发布代码和参数设置。

2. 相关作品

2.1. 多模式的个性化推荐

个性化推荐系统已成功应用于许多应用，如电子商务、新闻和社交媒体平台。通常，大多数现有的方法采用基于CF的方法（He, Du, Wang, Tian, Tang, Chua, 2018; He, Liao, Zhang, Nie, Hu, Chua, 2017; Rendle, 2010; Wang, He, Feng, Nie, Chua, 2018; Wang, He, Nie, Chua, 2017; Zhang, Shen, Liu, He, Luan, Chua, 2016; Tao, Wang, He, Huang, & Chua, 2019）。最近，随着深度神经网络（DNN）在计算机视觉、声学和自然语言处理任务中的成功（Hong et al., 2017; Hong, Yang, Wang, & Hua, 2015; Liu, Nie, Wang, Tian, & Chen, 2019; Nie et al., 2017; Wong, Chen, Mau, Sanderson, & Lovell, 2011），DNN也被引入到多模式领域（Nie et al., 2017; Wang et al., 2012）。特别是，一些人致力于将预先训练好的深度学习模型从多模态中提取的项目特征整合到基于CF的模型中，以增强项目表征。例如，*Chen*等人，Chen，He和Kan（2016）提出了一个名为CITING的模型，该模型挖掘并融合了文本特征，对社交媒体图片的语义进行建模，用于图片推文推荐。Covington、Adams和Sargin（2016）开发了一个由深度生成模型和排名模型组成的两阶段模型，用于视频推荐。Gao、Zhang和Xu（2017）开发了一个动态的递归神经网络，它融合了视频语义和用户兴趣来模拟用户的动态偏好。与这些方法不同的是，我们专注于对不同模式的用户偏好进行建模和拆分。

2.2. 图形卷积网络

由于其有效性和简单性，图卷积网络被广泛用于各种应用（van den Berg等人，2017；Hamilton, Ying, & Leskovec, 2017；Niepert, Ahmed, & Kutzkov, 2016；Perozzi, Al-Rfou, & Skiena, 2014）。通过图构操作，节点的局部结构信息可以通过按摩传递和聚合编码到它们的表示中。具体来说，PinSAGE（Hamilton等人，2017）是第一个已经成功应用于工业领域的基于GCN的模型，它通过采样和聚合其邻居的特征来生成节点的表示。同时，基于图的推荐方法也受到了很多关注（Cao, Wang, He, Hu, & Chua, 2019; Wang et al., 2019d; Wang, Jin, et al., 2020; Wang, Xu,2020 et al.）NGCF(Wang, He, Wang, et al.,)2019明确地将协作信号整合到嵌入过程中，并导致高阶特征交互在二元图中的表达式建模。LightGCN（He等人，2020）进一步简化了NGCF的设计，使其变得更容易训练并取得更好的性能。MMGCN（Wei等人，2019）试图在特定模型的用户-项目双字节图上对不同模式的用户偏好进行建模。

2.3. 门控注意力机制

为了控制信息传播的操作，门机制被引入到一些机器学习方案中。在LSTM（Hochreiter & Schmidhuber, 1997）中，该机制被用来实现输入门、遗忘门和输出门，它们分别用来记忆、遗忘和暴露记忆内容。 此外，在门机制的帮助下，GRU（Cho等人，2014）在递归网络中自适应地重置或更新其记忆内容。受这些方法中门机制的启发，一些基于图的模型利用该机制来提高其性能。例如，Li、Tarlow、Brockschmidt和Zemel（2016）设计了一个新的基于图的模型，被 称为GGCN，用来学习节点之间的长距离关系。与门控机制类似，注意力机制也被一些模块所采用，以权衡传播信息的重要性。AFM（Xiao等人，2017）采用了一个注意力网络来加权二阶交叉特征的重要性。Chen等人（2017）提出了一个注意力协作过滤框架，其中注意力被用来在两个层面上标记特征的重要性，以代表用户的偏好。考虑到图卷积运算中不同邻居的重要性，Velickovic等人（2017）设计了一个图关注网络，利用多头关注来控制信息传递。在我们的模型中，我们提出了一种门控注意力机制，利用门控和注意力机制的优势来控制和加重信息传播。

3. 任务制定

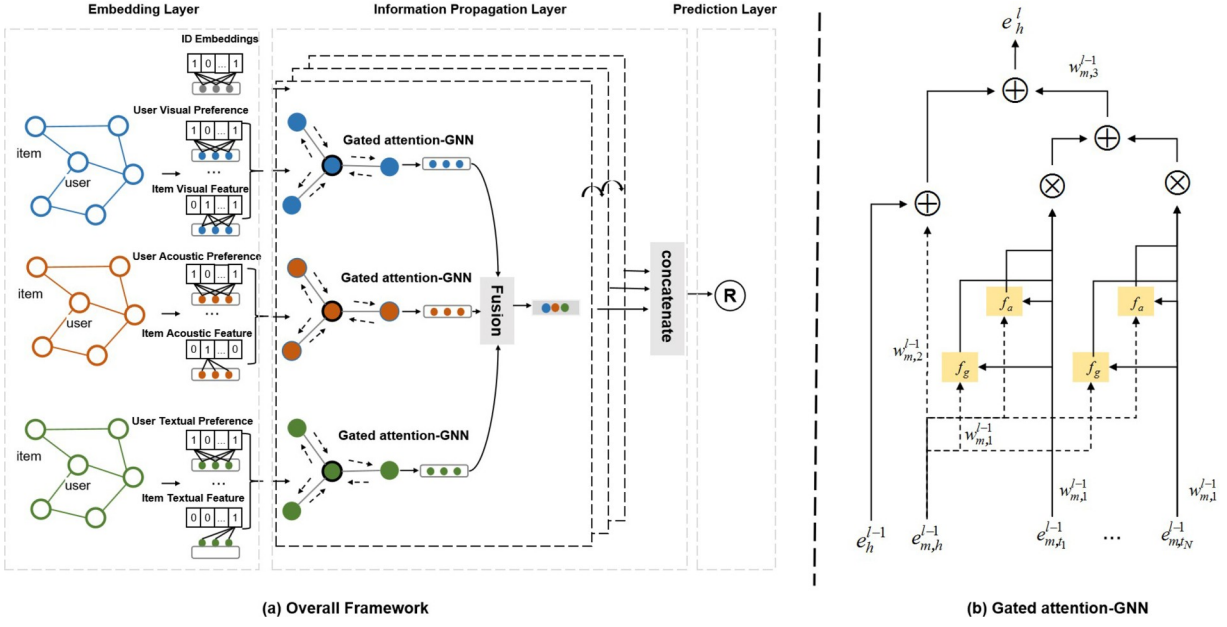
在这里，我们首先介绍我们的模型中使用的一些关键概念，然后再介绍任务的表述。

- **用户-物品二方图：**这种交互图建立在用户和物品的节点上，其中边是由用户和物品之间的历史交互构建的。我们将该图表述为$G = (V, E)$，其中$V = U \cup I$是涉及到的节点集。
  U和I分别作为用户和物品的集合，$E = \{(u, i)|u \in U, i \in I\}$是边集，每个边都代表用户$u$和物品$i$之间的相互作用。注意，我们将$G$形式化为一个无向图。
- **多模态交互图：**除了交互数据，一个多媒体项目（如微视频）通常与ID和多模态内容（*如*视觉、听觉和文本特征）相关联，而用户只是为了简单起见被分配了ID信息。对于每一种模式，在用户和项目节点之间设计了一个二方图，其中的边表示交互数据。更正式地说，我们将多模态交互图呈现为一个集合$\{G_m\}$，其中$m \in \{1, 2, 3\}$分别表示视觉、声音和文本模态。此外，出于用户隐私的考虑，我们只采用用户ID
- **高阶连通性：**受最近基于GNN的推荐方法的启发（Wang, He, Wang, et al., 2019; Wei et al, 2019），隐藏在用户行为数据中的用户-项目关系可以明确地表述为它们的连接性。特别是，用户的一阶连通性是由用户的历史项目组成的，直接剖析了他/她的偏好；类似地，项目的用户组可以作为描述性特征。此外，通过在图上进行随机行走，我们可以得到节点之间的高阶连接（*即*路径），这些连接编码了协作信号。以$i_1 \rightarrow u_1 \rightarrow i_2 \rightarrow u$的$_3$路径为例，我们可以得出结论，$u_1$和$u_3$很可能有相似的偏好，这是因为他们在采用$i$时有$_2$相似的行为；我们可以进一步将$u$选择的$i$推荐$_{11}$给$u_3$，这反映了协作过滤的效果。

- **任务描述。**我们现在将任务表述如下。
  - **输入。**多模态交互图，其每个节点都与视觉、声音和文本特征相关。
  - **输出。**预测用户u和物品i之间互动的可能性的推荐函数。

4. 方法

我们在此提出一个新的模型，名为多模态图注意网络（MGAT）。图1展示了MGAT的整体框架，它由四个部分组成。(1) 嵌入层，初始化用户和物品的ID嵌入；(2) 单模态交互图上的嵌入传播层，执行消息传递机制以捕捉用户对单个模态的偏好；(3) 跨多模态交互图的门控注意力聚合，利用与其他模态的相关性来学习每个邻居的权重，以指导传播；以及(4) 预测层，根据最终表示估计一个交互的可能性。



(a) Overall Framework　　(b) Gated attention-GNN

图1.左边是我们的模型框架，其中门控注意力机制被纳入信息传播过程，预测层的**R**是一个Sigmoid函数；右边是门控注意力-GNN结构的图示，其中f_a是注意力机制，f_g是门控机制。

4.1. 嵌入层

通常情况下，一个用户和一个物品都与ID有关。一个广泛使用的表示这种ID信息的解决方案是将ID嵌入为一个矢量表示。特别是，用户u和物品i被分别投影为$\mathbf{e}_u$和$\mathbf{e}_i$，这就记住了它们的一般特征。此外，在单个交互图中，每个物品i都有一个预先存在的特征$\mathbf{e}_{mi}$，以突出其在m-th模式中的特性。此外，我们为用户u分配了一个额外的嵌入$\mathbf{e}_{mu}$，以捕捉用户在m-th模式下的偏好。所有的嵌入都总结如下。

$$\mathbf{E} = \{\mathbf{e}_u, \mathbf{e}_i, \mathbf{e}_{m,u}, \mathbf{e}_{m,i} | u \in I, \quad i \in mM\} \tag{1}$$

其中，$\mathbf{e}_u$，$\mathbf{e}_{m,i} \in \mathbb{R}^{|U| \times d}$和$\mathbf{e}_i$，$\mathbf{e}_{m,i} \in \mathbb{R}^{|I| \times d}$；$N$和$M$分别表示用户和项目的数量；$d$是嵌入的大小。

值得注意的是，$\mathbf{e}_i$、$\mathbf{e}_u$和$\mathbf{e}_{m,u}$是在优化过程中随机初始化和训练的，而$\mathbf{e}_{m,i}$是通过可训练的神经网络从固定的特征中得到的。

4.2. 信息传播层

区别于直接将一对用户和项目嵌入到预测模型中的传统推荐模型，基于GNN的方法利用了交互图来细化表示。然而，对多模态交互图的研究还没有得到充分的探讨，现有的方法也未能在模态的粒度上分解用户的偏好。为此，我们将门控注意力机制引入到信息传播中。

4.2.1. 信息汇总

对于一个单独的交互图，考虑一个自我节点h与第一跳邻居$N_h = \{t | (h, t) \in E\}$，我们正式确定信息从这些邻居传播到$h$的情况如下。

$$e_h^{m,N} = \text{LeakyReLU}\left(\sum_{t \in N_h} f_a(h, t) f_g(h, t) e_{m,t} \mathbf{W} e_{m,t}\right) 。 \tag{2}$$

其中m是模态指标；$f_g(h, t)$和$f_a(h, t)$分别是门和注意成分在门控注意网络中的作用；$\mathbf{W}_{m,1}$是一个可训练的权重矩阵，用于提炼有用的线索。更具体地说，$f_g(h, t)$是传播门，用于决定信息是否会从t传播到h。同时，$f_a(h, t)$是注意力分数，表明t的贡献。

4.2.2. 传播门

受GRU（Cho等人，2014）最初采用的门机制的启发，之前的工作GGCN（Li等人，2016）采用了类似的组件来决定邻居是否会将信息传播给自我。在GGCN的启发下，我们利用这种门机制来控制执行传播时的信息流。我们使用以下三种类型的门来实现$f_g(h, t)$。

- *内积门*，首先计算$\mathbf{e}_{m,h}$和$\mathbf{e}_{m,t}$内积，并使用$\frac{1}{\sqrt{d}}$来处理不同的邻居数量。

  这被正式确定为

$$f_{gi}(h, t) = \delta\left(\frac{e_{m,h} e_{m,t}^{\top}}{\sqrt{d}}\right)。 \tag{3}$$

  其中$\delta(.)$是sigmoid函数，d是t的out degree。这种门取决于h和t之间的亲和力。*连接门*，连接两个表示，然后进行线性变换。

$$f_{gc}(h, t) = \delta\left(\frac{W_c(e_{m,h} || e_{m,t})}{\sqrt{d}}\right)。 \tag{4}$$

  其中||代表连接运算符，W_c是可训练的权重矩阵。

- 结合了这两种闸门并赋予这种机制以更大的灵活性的*双互动闸门*，其表述为：。

$$f_{gb}(h, t) = \delta\left(\frac{W_b(e_{m,h} || e_{m,t}) + e_{m,h} e_{m,t}}{\sqrt{d}}\right) \odot 。 \tag{5}$$

  其中，⊙是元素间的乘法运算。

因此，单模态交互图中的闸门分数反映了特定模态在剖析用户偏好时是否起作用。

4.2.3. 瞎邻友好的关注

此后，我们还引入了注意力机制来学习每个邻居的不同重要性，具体如下。

$$_{fa}(h, t) = (_{m,h}\mathbf{We}^{\top}_{m,h}) \tanh (W_{m,t}\mathbf{e}_{m,t}) 。 \quad (6)$$

其中tanh被用作非线性激活函数；而$W_{m,h}$和$W_{m,t}$是可学习的变换矩阵。为了简单起见，我们在这里考虑用内积来获得注意力权重，它反映了两个节点之间的亲和力。此后，我们采用softmax函数对所有邻居的注意力权重进行归一化，其表述如下。

$$f_a(h, t) = \frac{\exp _{fa}(h, t)}{\sum_{t' \in Nh} \exp _{fa}(h, t')}, \quad (7)$$

其中，最终的注意力分数能够区分出邻居的不同重要性分数。

在得到门和注意力的分数后，我们进行它们的乘积$f_g(h, t)f_a(h, t)$，*以便*在模式的颗粒度上传播个人兴趣。更具体地说，$f(_g h, t)$决定了个别模式中的项目是否会将信息传播给目标用户，而$f_a(h, t)$发现了这些项目对用户表征的不同贡献。

4.2.4. 信息组合

然后，我们利用从邻居$\mathbf{e}$传播的$_{m,Nh}$信息来更新节点h的表示。特别是，节点h的ID嵌入，$\mathbf{e}_h$，被视为跨模态的锚，作为执行跨模态propa-gation的高速公路。因此，我们首先将这个过程表述为：。

$$\tilde{\mathbf{e}}_{m,h} = \text{LeakyReLU}(_{m,2}\mathbf{We}_{m,h} + \mathbf{e}_h)。 \quad (8)$$

其中$\mathbf{W}_{m,2}$是转换矩阵；而$\mathbf{e}_h$实质上是连接$\{\mathbf{e}_{m,h}\}$的虚拟超节点，$\forall m \in M$。

$$\mathbf{e}^{(1)}_{m,h} = \text{LeakyReLU}\left(_{m,3}\mathbf{We}_{m,N} + \tilde{\mathbf{e}}_{m,h}\right)。 \quad (9)$$

其中$\mathbf{e}^{(1)}_{m,h}$表示编码一阶连接后的节点h的表示；$\mathbf{W}_{m,2}$是可训练的权重矩阵。

4.2.5. 高阶传播

按照之前的努力（Wang, He, Cao, et al., 2019; Wang, He, Wang, et al., 2019; Wei et al., 2019），我们可以叠加更多的信息传播层来利用节点之间的高阶连接，进一步丰富表示。更正式地说，节点h的表征被递归定义为。

$$\mathbf{e}^{(l)}_{m,h} = \text{LeakyReLU}\left(\mathbf{W}^{(l\,1)}_{m,3\ \mathrm{m},Nhm,h}\mathbf{e}^{(l\,1)} + \tilde{\mathbf{e}}^{(l\,))1}\right)。 \quad (10)$$

其中$\mathbf{e}^{(l\,1)}$是(l)之后的表示。 -1）传播步骤，存储来自（l -1）跳的邻居；而$\mathbf{e}^{(0)}$是$_{m,h}$的初始嵌入$\mathbf{e}_{m,h}$。

在更新了特定模态m中的节点表征后，我们可以将来自不同模态的表征组合成一个新的表征，可以表述为：。

$$e^{(l)}_h = \frac{1}{|M|}\sum_{m \in M} e^{(l)}_{m,h}. \quad (11)$$

4.3. 预测层

假设信息传播的数量为L，我们可以产生节点的最终表示，强调不同的邻居顺序，如下所示。

$$\mathbf{e}^*_u = \mathbf{e}^{(0)}_u \|\cdots\|_{\circ} \mathbf{e}^{(L)}_u \qquad \mathbf{e}^*_i = \mathbf{e}^{(0)}_i \|\cdots\|_{\circ} \mathbf{e}^{(L)}_i \quad (12)$$

最后，我们对用户和物品的表征进行内积，从而预测它们的匹配分数。

$$\hat{y}_{ui} = \mathbf{e}^{*}_u \mathbf{e}^{*}_i. \quad (13)$$

4.4. 优化

遵循主流的优化方法（He等人，2017；Rendle等人，2009；Wang，He，Wang等人，2019），我们采用贝叶斯个性化排名（BPR）来优化模型参数，它假设用户比那些没有事先互动的项目更喜欢以前互动过的项目。我们可以将其表述如下。

$$\mathcal{L} = \sum_{(u,i,j) \in O} - ln(\hat{y}_{uiuj}) + \lambda \|\theta\|^2_2 \quad (14)$$

其中，$O \in \{(u, i, j)|(u, i) \in R^+, (u, j) \in R^-\}$是训练数据集；$R^+$是包含用户$u$之间观察到的互动的数据集。

和项目 β_i；而 R 是未观察到的相互作用；$\delta(.)$ 是sigmoid函数；λ是衰减因子，θ 是模型中使用的参数。

5. 实验

在这一节中，我们详细介绍了我们的实验结果，其中包括MGAT的实验设置、性能比较和案例研究。

5.1. 实验设置

5.1.1. 数据集

为了评估MGAT的性能，我们在两个可公开获得的数据集上进行了实验，它们是MovieLens[1]，Tiktok[2]这两个数据集的统计数字见表，其详细情况1.,如下。

- **MovieLens数据集。**这个数据集已被广泛用于个性化推荐，它包含一系列的子集。在这项工作中，我们选择MovieLens-10M作为实验数据集。在多模态特征提取方面，我们采用了预先训练好的ResNet50（He, Zhang, Ren, & Sun, 2016）从视频的关键帧中提取视觉特征，声学特征是通过VGGish（Hershey等人，2017）从音频跟踪器中学习的。此外，文本特征由Sentence2Vector（Arora, Liang, & Ma, 2016）从文本内容中产生，其中包括标题和描述。
- **Tiktok数据集。**这个数据集是由流行的微视频分享平台Tiktok在一个数据挖掘竞赛中发布的。它包含了持续时间为3-15秒的微视频，以及用户提供的视频文字说明。我们使用了这个数据集中提供的原始脱敏的多模态特征向量。所有的模式特征都是脱敏的，并以矢量的方式提供，没有原始数据。这些视频的持续时间为3-15秒，文本内容来自用户提供的说明。

我们将每个数据集随机分为三部分--训练（80%）、验证（10%）和测试（10%）集。验证集被用来调整超参数，我们选择最佳训练模型。我们报告测试集上表现最好的模型的最终性能。

5.1.2. 评价指标

本节介绍了我们实验中使用的评价指标和参数设置。我们采用了三个广泛使用的评价指标。Precision@K, Recall@K, 和NDCG@K。我们设定K=，并10报告测试集中所有用户取得的平均性能。每个用户的负面项目被定义为与该用户没有互动的项目。所有实验的代码都是用PyTorch框架实现的。对于所有的模型，嵌入大小是在64所有模型中，批处理大小是1024。我们采用Xavier初始化器来初始化所有的模型参数。此后，我们用Adam优化器优化所有模型。此外，我们应用网格搜索来进行超参数细化，其中学习率的值从 $\{1e-1, 1e-2, 1e-3, 1e-4, 1e-5\}$，而那些用于权重衰减和注意力下降的比率则选自 $\{1e-1, 1e-2, 1e-3, 1e-4, 1e-5\}$，$\{0.1, .0.2, .., 0.8\}$，分别。在没有说明的情况下，节点掉线和消息掉线是 0.0.其他基线使用原始论文中使用的超参数。

5.1.3. 基线

为了评估我们模型的性能，我们将MGAT与以下基线进行比较。

- **VBPR**（He & McAuley，2016）。这个模型将视觉特征注入到物品的表示中。随后，它利用矩阵分解法，根据用户和物品的历史互动，学习它们的表征。在我们的实验中，我们将微视频的多模态特征串联成一个特征向量。然后，我们将其与ID信息整合，预测用户和物品之间的互动。
- **ACF**（Chen等人，2017）。它引入了项目级和组件级注意力来处理多媒体推荐的隐性反馈。在我们的实验中，我们对每个模式的交互预测采用了类似的组件级关注机制。
- **GraphSAGE**（Hamilton等人，2017）。它是一个基于图的模型，聚合了来自邻居节点的信息来表示未见过的数据。在这项工作中，我们将三种模式的特征串联起来以表示每个节点。
- **NGCF** (Wang, He, Wang, et al., 2019)。NGCF以明确的方式将协作信号整合到嵌入过程中。它通过纳入来自多级邻居的信息传递，对双胞胎图中的高阶特征互动进行建模。在本文中，我们将所有的多模态特征串联起来作为项目表示，在NGCF中用于嵌入交互的次序过程。
- **MMGCN**（Wei等人，2019）。MMGCN是一种基于图的算法。为了学习用户对不同模式的偏好，它根据每种模式的用户与物品的互动，设计了一个特定模式的双方图。之后，它将所有特定模型的表征集合起来，以获得用于预测的用户或项目的表征。

1 https://grouplens.org/datasets/movielens/。

[2]http://ai-lab-challenge.bytedance.com/tce/vc/。

表 1
Tiktok和MovieLens数据集的统计数据。注意，V、A和T的符号分别代表用于原始视觉、声音和文本数据的特征数量。

数据集	#互动	#Items	#Users	稀缺性	V	A	T
Tiktok	726,065	76,085	36,656	99.99%	128	128	128
电影	1,239,508	598,655	5,000	99.63 100			2048

5.2. 性能比较

所有的实验结果如表2所示，我们有以下观察。

- MGAT的表现优于所有的基线模型。这证明了我们模型的合理设计。与传统的只考虑用户-项目直接联系的结点过滤方法（VBPR和ACF）相比，我们的MGAT使用高阶连接来促进表示学习。与基于GNN的推荐器（GraphSage和NGCF）相比，MGAT使用一个双点图，并简单地将不同模式的特征统一为一个，MGAT识别了三个通道来传播有用的信号，具有更好的表示能力。与MMGCN应用后融合来整合各个模态的表征相比，我们的MGAT通过精心设计的注意力机制来区分各个模态的重要性，从而识别用户的细粒度偏好。
- 两种基于多模态图的模型都超过了其他基线的表现。与从平等加权的多模态特征中学习偏好的算法相比，MGAT和MMGCN始终取得更好的性能。这表明，对不同模态的关注有助于更好地建立用户偏好模型。
- 在Tiktok数据集上，基于图的模型的表现优于基于CF的模型。它验证了通过消息传递将邻居的信息注入到节点表示中可以改善微视频的表示。此外，GraphSAGE和NGCF在MoviLens上的表现比VBPR要差。我们把这种发现归结为数据的差异：由于Tiktok数据集的视频比MovieLens的短，长视频的原始特征会包含更多的复杂信息甚至是噪音，而且多种模式之间的关系是高度纠结的。
- 出乎意料的是，ACF在所有实验中的表现都很差。这可能是由于我们修改了ACF算法的实现，为了公平比较，我们用组件级的特定模态特征代替了这些特征。

5.3. MGAT的案例研究

在本节中，我们将介绍MGAT的案例研究，以调查可能对我们的模型性能产生影响的因素。

5.3.1. 高阶连通性的影响

在这一节中，我们评估了高阶连接是如何影响MGAT的性能的。具体来说，我们对MGAT的三种变体进行了体验，这些变体采用了不同顺序的邻居来表示节点。例如，MGAT_3意味着使用三阶邻居。此外，嵌入的大小在64两个数据集上都有。

如表所示3。对不同特征顺序进行建模的MGAT变体之间的性能比较。根据我们的实验，MGAT_2比MGAT_1和MGAT_3表现更好，这与NGCF（Wang, He, Wang, et al., 2019）、MMGCN（Wei et al., 2019）、GAT（Velickovic et al., 2017）论文中的观察一致。这表明图神经网络的过度平滑问题，即堆叠更多的图卷积层或从高阶邻居中提炼信号，容易引入远程邻居的噪声，导致次优性能。

5.3.2. 门和注意力的影响

在我们的模型中，我们引入了门控注意力机制来控制和加重信息的传播。为了研究它对模型的影响，我们在两个数据集上进行了实验，以评估二阶MGAT模型在不同的情况下的性能。

表 2

MGAT与最先进的推荐算法在Tiktok和MoviewLens数据集上的性能比较。

	Tiktok			MovieLens		
Model	Precision	Recall	NDCG	Precision	Recall	NDCG
VBPR	0.0972	0.4878	0.3136	0.1172	0.4724	0.2852
ACF						

表 1

0.08730.44290.28670.10780.4304 0.2589

图谱

0.10280.49720.32100.11320.4532 0.2647

NGCF

0.10650.50080.32260.11560.4626 0.2732

MMGCN

0.11640.55200.34230.12110.5138 **0.3062**

MGAT

0.12510.59650.38380.12720.5412 **0.3251**

增长率%。

7.47%8.06%12.12%5.03%5.33%6.17%

表 3
不同顺序的影响。

	TiktokMovieLens					
	ModelPrecisionRecallNDCGPrecisionRecallNDCG					
MGAT_1	0.1210.57730.36810.12130.512	0.3054				
MGAT_2	**0.12510.59650.38380.12720.5412**	**0.3251**				
MGAT_3	0.12030.57750.36570.12420.5262	0.3145				

门控和注意机制的组合。MGAT_no表示没有注意和门控机制的基础模型，MGAT_g表示没有门控机制的基础模型，MGAT表示有门控注意机制的基础模型，结果见表4.

如表4所示，MGAT_g的性能优于MGAT_no，这验证了通过引入门机制来控制来自邻居的传播信息所带来的性能提升。同时，MGAT的性能优于MGAT_g，这说明门控注意力机制可以通过对传播信息的加权来提高模型的性能。此外，MGAT_g是只抛弃了注意力机制，而保留了门控机制的变体。与MGAT_g相比，MGAT取得了更好的性能，表明注意力机制具有积极的作用。

5.3.3. 不同门机制的影响

为了评估不同的门控注意力机制引起的不同效果，我们对门控注意力机制的三种变体进行了实验，包括内部传播门（MGAT_i）、串联传播门（MGAT_c）和Bi-interaction传播门（MGAT_bi）。

如表5所示，MGAT_i的性能优于其他两个模型。这表明内积可能更适合于在我们基于多模态图的模型中对信息关系进行建模。相比之下，其他带有转换矩阵的模型可能会因为过度参数化的转换矩阵而出现过拟合现象。

5.4. 一个案例研究

正如第4节所介绍的，我们的模型采用了门控注意力机制，它被用来控制和权衡形成中的流量传播。在图2中，我们对某个节点的10个邻居进行了抽样调查，然后将节点的门控注意力机制值可视化。行中表示邻居的索引，列中表示参数的注意值。此外，颜色代表不同模式的值。

如图所示2，我们观察到不同的节点在模式方面有不同的由门控注意力机制产生的注意力值。这表明，邻居的重要性是不同的。此外，在大多数节点中，不同模态的注意值也是不同的；这表明不同模态的特征对同一节点的重要性也是不同的。此外，这种机制使MGAT具有更好的可解释性，这也是推荐系统中普遍关注的问题（Ren, Liang, Li, Wang, & de Rijke, 2017）。

6. 结论

本文提出了一种基于图的算法，名为MGAT，该算法通过高阶邻接内形成和不同模态的注意机制来模拟用户的偏好。具体来说，在对用户偏好进行建模的过程中，通过高阶连接和消息传递机制将多模态特征注入到嵌入中。同时，引入了门控注意力机制来控制和加权来自高跳邻居的传播的信息。在两个真实世界的数据集上进行的广泛实验证明了我们模型的有效性。在我们的模型中，我们对用户在不同模式上的预判进行建模，而微视频中有更多类型的特征，如关系（Shang等人，2019）和因果关系。因此，我们将更加关注对微视频的理解，并在未来的工作中利用更多的特征进行微视频的再赞扬。

表 4

门和注意力机制的影响。

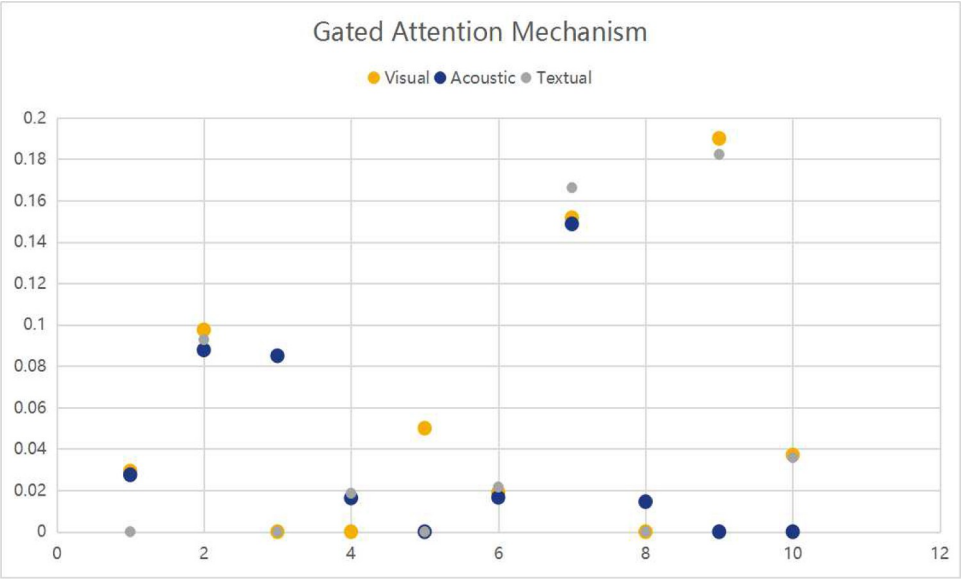
	TiktokMovieLens					
	ModelPrecisionRecallNDCGPrecisionRecallNDCG					
MGAT_no						

MGAT_g	0.11980.56440.36210.12510.5323	0.3187
MGAT_g	0.12150.57480.36290.1260.5344	0.319
MGAT	**0.12510.59650.38380.12720.5412**	**0.3251**

表3

表 5
不同门机制的影响。

	TiktokMovieLens					
	Model	Precision	Recall	NDCG	Precision	Recall	NDCG
MGAT_i	MGAT_c	0.1224	0.5866	0.3753	0.124	0.5263	0.3153
	MGAT_i	**0.1251**	**0.5965**	**0.3838**	**0.1272**	**0.5412**	**0.3251**
MGAT_bi		0.1222	0.5887	0.3773	0.1249	0.527	0.3162



图：来自Tiktok数据集的门控注意力机制的可视化**2.**。

CRediT作者的贡献声明

陶竹林。写作-原稿，可视化，调查，验证，调查，软件，方法学，概念化。**魏银伟。**资源，数据整理，写作-审查和编辑，监督。**王翔。概念化**，写作-审查和编辑，监督。**何湘南：写作-审查和编辑。**写作-审查和编辑。**黄祥林：**写作-审查和编辑。**蔡达生：写作-审查和编辑。**写作-审查和编辑。

鸣谢

这项研究是NExT++研究的一部分，也得到了新加坡国家研究基金会在其AI新加坡计划下、Linksure Network Holding Pte Ltd和亚洲大数据协会的支持（奖励编号：AISG-100E-2018-002）。NExT++由新加坡总理办公室国家研究基金会在其IRC@SG资助计划下支持。此外，本研究还得到中国国家重点研发计划（No.2019YFB1406201）和未来学校计划（No.CSDP17FS3231）的支持。

补充材料

与本文相关的补充材料可在网上版本中找到，网址是10.1016/j.ipm.2020.102277。

参考文献

Arora, S., Liang, Y., & Ma, T. (2016). *一个简单但难以战胜的句子嵌入基线。* ICLR1-16。
Tao, Z., Wang, X., He, X., Huang, X., & Chua, T. (2019).HoAFM: 用于CTR预测的高阶注意力因式分解机.*Information Processing&Management,*

102076.

van den Berg, R., Kipf, T. N., & Welling, M. (2017).图卷积矩阵完成。 *CoRR， abs/1706.02263*。

Cao, Y., Wang, X., He, X., Hu, Z., & Chua, T.（2019）.*统一知识图谱学习和推荐。实现对用户偏好的更好理解。* WWW151-161.

Chen, J., Zhang, H., He, X., Nie, L., Liu, W., & Chua, T. (2017).*注意力协作过滤。具有项目和组件级注意力的多媒体推荐。* SIGIR335-344。

Chen, T., He, X., & Kan, M. (2016).*上下文感知的图像推文建模和推荐*acm1018-1027.

Chen, T., Zhang, W., Lu, Q., Chen, K., Zheng, Z., & Yu, Y. (2012).Svdfeature:基于特征的协同过滤的工具箱。*JMLR,* 3619-3622.

Cho, K., van Merrienboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014).*使用RNN编码器-解码器为统计机器翻译学习短语代表。* emnlp1724-1734.

Covington, P., Adams, J., & Sargin, E. (2016).*用于YouTube推荐的深度神经网络。* RecSys191-198。

Fan, W., Ma, Y., Li, Q., He, Y., Zhao, Y. E., Tang, J., & Yin, D. (2019).*图神经网络的社交推荐。* WWW417-426.Gao, J., Zhang, T., & Xu, C. (2017).*A unified personalized video recommendation via dynamic recurrent neural networks.ACM* MM127-135.Hamilton, W., Ying, Z., & Leskovec, J. (2017).*大型图上的归纳表征学习。* NIPS1024-1034.

He, K., Zhang, X., Ren, S., & Sun, J. (2016).*用于图像识别的深度残差学习。* CVPR770-778.He, R., & McAuley, J. (2016).*Vbpr:Visual Bayesian personalized ranking from implicit feedback.*AAAI1-8.He, X., & Chua, T. (2017).*Neural factorization machines for sparse predictive analytics.*SIGIR355-364.

He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., & Wang, M. (2020).Lightgcn:简化和增强图卷积网络的推荐功能。*CoRR, abs/. 2002.02126*.

He, X., Du, X., Wang, X., Tian, F., Tang, J., & Chua, T. (2018).*基于外部产品的神经协作过滤.*ijcai2227-2233.

He, X., He, Z., Song, J., Liu, Z., Jiang, Y., & Chua, T.（2018）。NAIS：用于推荐的神经entive项目相似度模型。*TKDE,30* (12), 2354-2366.He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. （2017）.*神经协作过滤。* WWW173-182.

Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., Seybold..., B., et al. （2017）。*用于大规模音频分类的CNN架构。* *ICASSP*. IEEE131-135.

Hochreiter, S., & Schmidhuber, J. (1997).长短期记忆。*神经计算，9*（8），1735-1780。

Hong, R., Li, L., Cai, J., Tao, D., Wang, M., & Tian, Q（2017）。连贯的语义-视觉索引用于云中的大规模图像检索。*IEEE Transactions on Image Processing,26* (9), 4128-4138.

Hong, R., Yang, Y., Wang, M., & Hua, X. (2015).学习视觉语义关系以实现高效的视觉检索。*IEEE Transactions on Big Data,1* (4), 152-161.Kabbur, S., Ning, X., & Karypis, G. (2013).*FISM：用于top-n推荐系统的因子项目相似性模型。* SIGKDD659-667.

Koren, Y., Bell, R. M., & Volinsky, C. (2009).推荐系统的矩阵分解技术。*IEEE Computer,* 30-37.Li, Y., Tarlow, D., Brockschmidt, M., & Zemel, R. S. (2016).*Gated graph sequence neural networks.ICLR*.

Liu, M., Nie, L., Wang, X., Tian, Q., & Chen, B. (2019).在线数据组织者。通过结构引导的多模态字典学习对微视频进行分类。*IEEE Transactions on Image Processing,28* (3), 1235-1247.

Nie, L., Wang, X., Zhang, J., He, X., Zhang, H., Hong, R., & Tian, Q.（2017）。*通过利用外部声音加强微视频的理解。* acm1192-1200.Niepert, M., Ahmed, M., & Kutzkov, K. (2016).*学习图的卷积神经网络。* ICML2014-2023。

Ning, X., & Karypis, G. (2011).*SLIM: 顶层推荐系统的稀疏线性方法.*ICDM497-506.Perozzi, B., Al-Rfou, R., & Skiena, S. (2014).*Deepwalk:社会表征的在线学习。* SIGKDD701-710。

Ren, Z., Liang, S., Li, P., Wang, S., & de Rijke, M. (2017).*具有可解释建议的社会协作观点回归。* WSDM485-494.Rendle, S. (2010).*因式分解机* .ICDM995-1000.

Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Tieme, L. (2009).*Bpr：来自隐性反馈的贝叶斯个性化排名。* UAI452-461.Shang, X., Di, D., Xiao, J., Cao, Y., Yang, X., & Chua, T. (2019).*注释用户生成的视频中的对象和关系。* ICMR279-287.

Song, W., Xiao, Z., Wang, Y., Charlin, L., Zhang, M., & Tang, J.（2019）。*通过动态图注意网络进行基于会话的社交推荐。* WSDM555-563.Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. （2017）.*图形注意力网络。*

Wang, H., Zhao, M., Xie, X., Li, W., & Guo, M. (2019).*用于推荐系统的知识图谱卷积网络.*WWW3307-3313.

Wang, M., Hong, R., Li, G., Zha, Z., Yan, S., & Chua, T. (2012).通过标签定位和关键镜头识别进行事件驱动的网络视频总结。*IEEE Transactions on Multimedia,14* (4), 975-985.

Wang, X., He, X., Cao, Y., Liu, M., & Chua, T. (2019).*KGAT：用于推荐的知识图谱关注网络。* KDD950-958.

Wang, X., He, X., Feng, F., Nie, L., & Chua, T. (2018).*TEM：用于可解释推荐的树状增强嵌入模型。* WWW1543-1552.Wang, X., He, X., Nie, L., & Chua, T. (2017).*项目丝绸之路。从信息领域向社会用户推荐项目。* SIGIR185-194.

Wang, X., He, X., Wang, M., Feng, F., & Chua, T. (2019).*神经图协作过滤。* SIGIR165-174.

Wang, X., Jin, H., Zhang, A., He, X., Xu, T., & Chua, T. (2020).Disentangled Graph Collaborative Filtering.*SIGIR*.

Wang, X., Wang, D., Xu, C., He, X., Cao, Y., & Chua, T. (Wang, Xu, He, Cao, Chua, 2019d).*用于推荐的知识图谱上的可解释推理。* *AAAI*.Wang, X., Xu, Y., He, X., Cao, Y., Wang, M., & Chua, T. (2020).Reinforced Negative Sampling over Knowledge Graph for Recommendation.*WWW,* 99-109.

Wei, Y., Wang, X., Nie, L., He, X., Hong, R., & Chua, T. (2019).*MMGCN：用于微视频个性化推荐的多模态图卷积网络。* acm1437-1445.

Wong, Y., Chen, S., Mau, S., Sanderson, C., & Lovell, B. C. (2011).*基于补丁的概率图像质量评估，用于人脸选择和改进基于视频的人脸识别。* CVPR74-81.

Wu, L., Sun, P., Fu, Y., Hong, R., Wang, X., & Wang, M. (2019).*社会推荐的神经影响扩散模型。* SIGIR235-244.

Wu, Q., Zhang, H., Gao, X., He, P., Weng, P., Gao, H., & Chen, G.（2019）。*双图注意网络用于推荐系统中多面社会效应的深度潜伏表示。* WWW2091-2102.

Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X., & Tan, T.（Tang, Zhu, Wang, Xie, Tan, 2019c）。*基于会话的图神经网络推荐。* AAAI346-353.Xiao, J., Ye, H., He, X., Zhang, H., Wu, F., & Chua, T. (2017).*注意力分解机。通过注意力网络学习特征相互作用的权重。* ijcai3119-3125.

Zhang, H., Shen, F., Liu, W., He, X., Luan, H., & Chua, T. (2016).*离散协同过滤.*SIGIR325-334.Zheng, L., Lu, C., Jiang, F., Zhang, J., & Yu, P. S. (2018).*Spectral collaborative filtering.*RecSys311-319.

Zheng, Y., Gao, C., He, X., Li, Y., & Jin, D. (2020).图卷积网络的价格感知推荐.*CoRR, abs/2003.03975*.