

基于图神经网络的推荐系统研究

毕业设计结题答辩

学生：周华鹏 指导老师：张毅锋

东南大学信息科学与工程学院

2023 年 6 月 6 日



① 课题背景

② 基于 GAT 的推荐系统

③ 实验研究

④ 参考文献

① 课题背景

本文研究内容

图和图神经网络

GraphRec

② 基于 GAT 的推荐系统

③ 实验研究

④ 参考文献

① 课题背景

本文研究内容

图和图神经网络

GraphRec

② 基于 GAT 的推荐系统

③ 实验研究

④ 参考文献

工作概览

- 提出了三种模型，完成了预处理数据集到对比多组实验等。
- 引入物品的种类信息，提出 GraphRC，在评价指标上比基线模型好了大约 11%.
- 引入交叉注意力机制，提出 GraphRCC，在评价指标上比基线模型好了大约 16%.
- 所有主体实验代码、用到的数据集已在 github 开源：<https://github.com/HuapengZhou/GraphRecPlus>

① 课题背景

本文研究内容

图和图神经网络

GraphRec

② 基于 GAT 的推荐系统

③ 实验研究

④ 参考文献

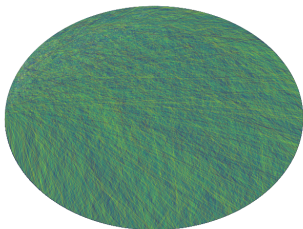
背景介绍

- 图广泛存在在我们的生活中。
- 图代表的点边关系较为复杂, 在推荐系统中可以用图以及图神经网络处理。
- 用 GCN 处理能够获取邻居的信息。¹
- 用 GAT 处理邻居节点可以给到一个加权系数。²

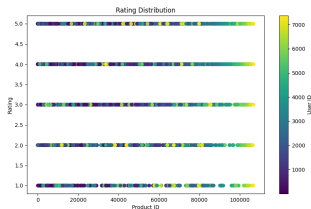
¹[1]T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in International Conference on Learning Representations, 2017.

²[2]P. Velickovi, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," arXiv preprint arXiv:1710.10903, 2017.

处理任务的可视化



(a)



(b)

图 1: 用户直接的关系以及对物品的评价信息

GCN

- 图卷积神经网络将卷积操作从图像处理领域扩展到了图结构数据中。
- 图卷积网络（GCN）的工作机制主要是基于以下的公式：

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (1)$$

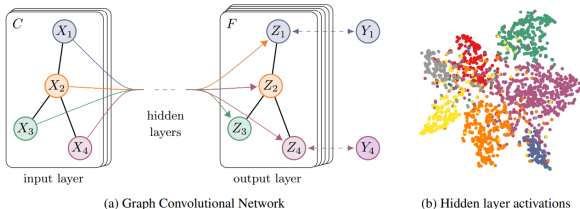


图 2: GCN

GAT

- GAT 借鉴了 Transformer 的注意力机制，引入了 masked self-attention 机制。

(GCN) 的工作机制主要是基于以下的公式：

$$\alpha_{ij} = \frac{\exp \left(\text{LeakyReLU} \left(\vec{\mathbf{a}}^T \left[\mathbf{W} \vec{h}_i \| \mathbf{W} \vec{h}_j \right] \right) \right)}{\sum_{k \in \mathcal{N}_i} \exp \left(\text{LeakyReLU} \left(\vec{\mathbf{a}}^T \left[\mathbf{W} \vec{h}_i \| \mathbf{W} \vec{h}_k \right] \right) \right)} \quad (2)$$

- 我们就可以对邻居节点的特征进行加权平均，得到节点 i 的新特征 \vec{h}'_i ：

$$\vec{h}'_i = \sigma \left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{N}_i} \alpha_{ij}^k \mathbf{W}^k \vec{h}_j \right) \quad (3)$$

① 课题背景

本文研究内容

图和图神经网络

GraphRec

② 基于 GAT 的推荐系统

③ 实验研究

④ 参考文献

GraphRec 结构

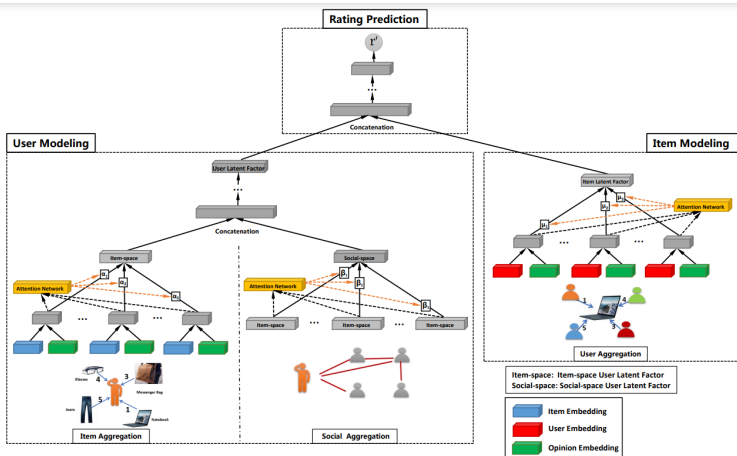


图 3: GraphRec 结构

问题

- 网络模型³不够先进。深层网络会过拟合现象比较严重。
- 没有考虑用户对物品种类的评价信息。
- 没有考虑其他信息对加权因子的影响。

³[3]W. Fan and et al., “Graph neural networks for social recommendation,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 7370–7377.

1 课题背景

2 基于 GAT 的推荐系统

所提出的模型总览

GraphRes

GraphRC

GraphRCC

模型训练

3 实验研究

4 参考文献

1 课题背景

2 基于 GAT 的推荐系统

所提出的模型总览

GraphRes

GraphRC

GraphRCC

模型训练

3 实验研究

4 参考文献

总览图

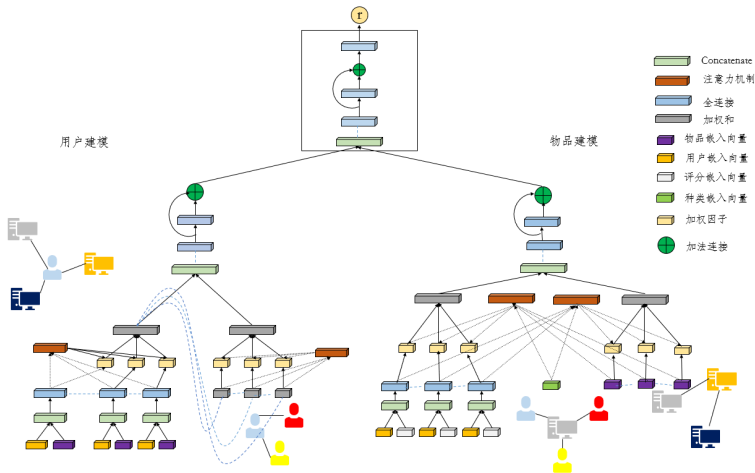


图 4: 所提出的模型图

1 课题背景

2 基于 GAT 的推荐系统

所提出的模型总览

GraphRes

GraphRC

GraphRCC

模型训练

3 实验研究

4 参考文献

ResNet

- ResNet[4] 是为了解决深度神经网络的退化问题而设计的, 阻止了梯度消失, 利用了深度网络的强大的数据提取能力, 公式如下:

$$x_{l+1} = F(x_l) + x_l \quad (4)$$

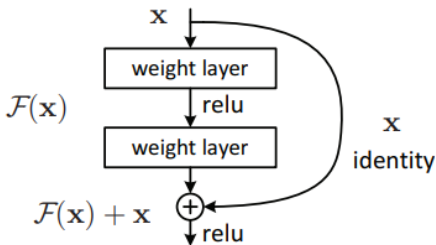


图 5: 残差连接

添加残差块

- 在用户建模, 物品建模, 评级预测中都添加残差块, 以评级预测举例:

$$\begin{aligned}
 \mathbf{g}_1 &= [\mathbf{h}_i \oplus \mathbf{z}_j] \\
 \mathbf{g}_2 &= \sigma(\mathbf{W}_2 \cdot \mathbf{g}_1 + \mathbf{b}_2) + \mathbf{g}_1 \\
 &\dots \\
 \mathbf{g}_l &= \sigma(\mathbf{W}_l \cdot \mathbf{g}_{l-1} + \mathbf{b}_l) + \mathbf{g}_{l-1} \\
 r'_{ij} &= \mathbf{w}^T \cdot \mathbf{g}_l
 \end{aligned} \tag{5}$$

残差块展示

- 添加了以下的残差块, 便形成了模型 1, GraphRes:

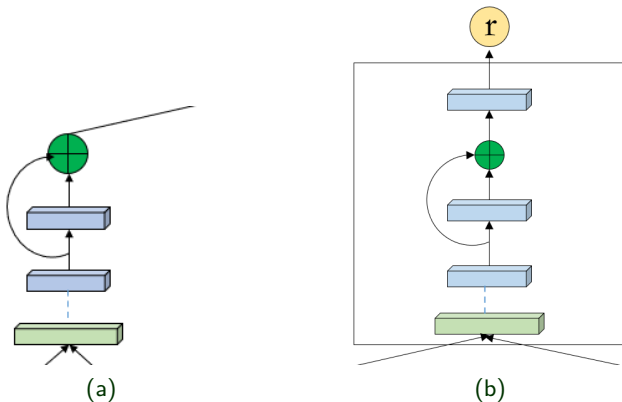


图 6: GraphRes 用到的残差块展示

1 课题背景

2 基于 GAT 的推荐系统

所提出的模型总览

GraphRes

GraphRC

GraphRCC

模型训练

3 实验研究

4 参考文献

GraphRC 建模

将物品 v_j 的种类嵌入向量 k_w 加入到前面的公式中, 得到新的交互用户表示 f_{jt} :

$$f_{jt} = g_u ([p_t \oplus e_r \oplus k_w]) \quad (6)$$

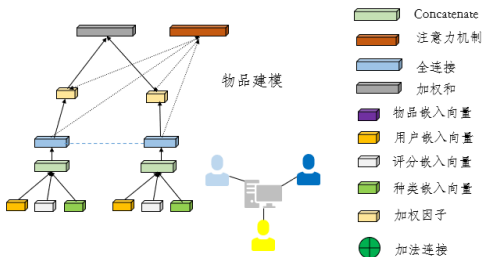


图 7: 所提出算法的 GraphRC 的物品建模

1 课题背景

2 基于 GAT 的推荐系统

所提出的模型总览

GraphRes

GraphRC

GraphRCC

模型训练

3 实验研究

4 参考文献

交叉注意力

交叉注意力是一种能够捕捉两个不同序列之间交互关系的注意力机制。[5]

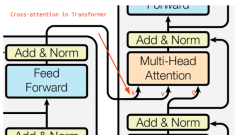


图 8: 交叉注意力机制

注意力机制中，查询 (Q)，键 (K) 和值 (V) 的关系通常被表达为以下的公式：

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (7)$$

加入交叉注意力机制

以物品建模中的用户聚合为例：

1. 查询 (Q) 是 f_{jt} ，即用户 t 对物品 j 的交互表示。
 2. 键 (K) 是种类的嵌入向量 \mathbf{k}_w 。
 3. 值 (V) 是 \mathbf{q}_i ，即其他同时属于种类 w 的物品 i 的嵌入向量。
- 首先，我们可以计算注意力因子 μ_{jt} ：

$$\mu_{jt} = \frac{\exp(\mathbf{f}_{jt}^T \mathbf{k}_w \mathbf{q}_i)}{\sum_{i' \in D(j)} \exp(\mathbf{f}_{jt}^T \mathbf{k}_w \mathbf{q}_{i'})} \quad (8)$$

其中， D 是和物品 j 同时属于种类 w 的物品集合。

在物品建模中的物品聚合中可以进行同样的操作，最后拼接出来形成新的物品隐藏因子建模

GraphRCC 效果图

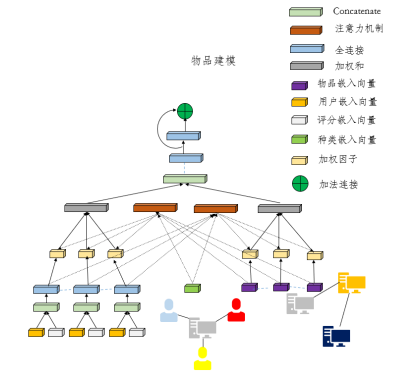


图 9: 最终的物品建模

1 课题背景

2 基于 GAT 的推荐系统

所提出的模型总览

GraphRes

GraphRC

GraphRCC

模型训练

3 实验研究

4 参考文献

模型训练情况

所选取的 loss 函数为均方误差，具体公式如下：

$$\text{Loss} = \frac{1}{2|O|} \sum_{i,j \in O} (r'_{ij} - r_{ij})^2 \quad (9)$$

其中， $|O|$ 表示观察到的评分数量， r_{ij} 是用户 i 对物品 j 的真实评分。

模型训练时采用 dropout 策略防止过拟合等。[6]。

① 课题背景

② 基于 GAT 的推荐系统

③ 实验研究

实验设置

实验结果

④ 参考文献

① 课题背景

② 基于 GAT 的推荐系统

③ 实验研究

实验设置

实验结果

④ 参考文献

数据集

- 数据集：购物数据集 Ciao (<https://www.ciao.co.uk/>)

表 1: Ciao 数据集

用户数目	7,317
项目数目	10,4975
评分数目	283,319
评分密度	0.0368 %
社交连接数目	111,781
社交关系密度	0.2087 %

评价指标

- 评价指标：平均绝对误差（Mean Absolute Error, MAE）和均方根误差（Root Mean Square Error, RMSE）

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (11)$$

① 课题背景

② 基于 GAT 的推荐系统

③ 实验研究

实验设置

实验结果

④ 参考文献

超参数分析

表 2: 相同模型不同参数的结果对比

Model	RMSE	MAE	DataSet(Subset)	Dim	1r
GraphRes	1.1618	0.9851	Train:0.01 Test:0.1	8	0.001
GraphRes	0.7735	0.5574	Train:0.01 Test:0.1	64	0.001
GraphRes	0.8191	0.6280	Train:0.01 Test:0.1	8	0.01
GraphRes	0.8128	0.5824	Train:0.03 Test:0.1	8	0.001
GraphRes	0.8105	0.5374	Train:0.07 Test:0.1	8	0.001
GraphRes	0.8013	0.5291	Train:0.1 Test:0.1	8	0.001
GraphRes	0.7801	0.5595	Train:0.1 Test:1	8	0.001
GraphRes	0.7236	0.4792	Train:1 Test:1(3 epoch)	8	0.001

实验结果

评价指标及损失函数曲线

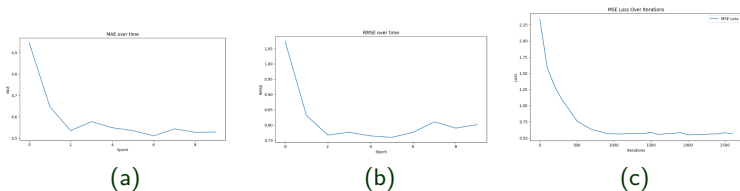


图 10: 第七组实验的评价指标及 loss 曲线

GraphRes 和 GraphRec 多组消融实验

表 3: 相同参数不同模型的结果对比

模型	RMSE	MAE	DataSet	Dim	1r
GraphRes	1.1618	0.9851	Train:0.01 Test:0.1	8	0.001
GraphRec	1.1838	1.0260	Train:0.01 Test:0.1	8	0.001
GraphRes	0.7801	0.5595	Train:0.1 Test:1	8	0.001
GraphRec	0.8967	0.7268	Train:0.1 Test:1	8	0.001
GraphRes	0.7735	0.5574	Train:0.01 Test:0.1	64	0.001
GraphRec	0.9720	0.7661	Train:0.01 Test:0.1	64	0.001

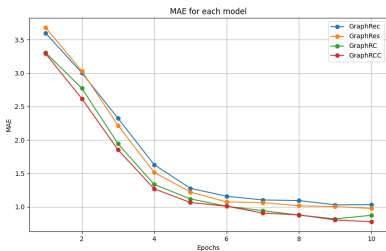
四种模型消融实验

表 4: GraphRec, GraphRes, GraphRC 和 GraphRCC 四种模型每个 epoch 不同的 RMSE 和 MAE 值

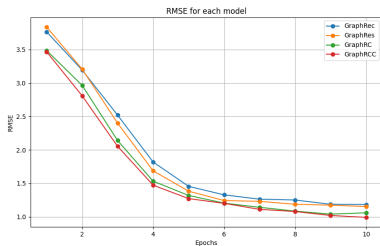
轮数	GraphRec	GraphRes	GraphRC	GraphRCC
1	3.7664/3.6004	3.8384/3.6834	3.4868/3.3063	3.4684/3.2934
2	3.1950/3.0060	3.2088/3.0317	2.9663/2.7752	2.8097/2.6193
3	2.5199/2.3245	2.3998/2.2180	2.1403/1.9445	2.0528/1.8532
4	1.8189/1.6305	1.6872/1.5149	1.5323/1.3346	1.4726/1.2688
5	1.4552/1.2767	1.3829/1.2202	1.3190/1.1181	1.2717/1.0643
6	1.3275/1.1580	1.2422/1.0727	1.2041/ 1.0086	1.2004/1.0087
7	1.2635/1.1022	1.2309/1.0631	1.1415/0.9419	1.1106/0.9062
8	1.2506/1.0940	1.1867/1.0154	1.0845/ 0.8756	1.0783/0.8790
9	1.1869/1.0289	1.1737/1.0033	1.0392/0.8188	1.0192/0.8036
10	1.1829/1.0319	1.1525/0.9755	1.0594/0.8728	0.9911/0.7765

实验结果

四种模型评价指标对比图



(a)



(b)

图 11: 对比图

实验结果

运行时间

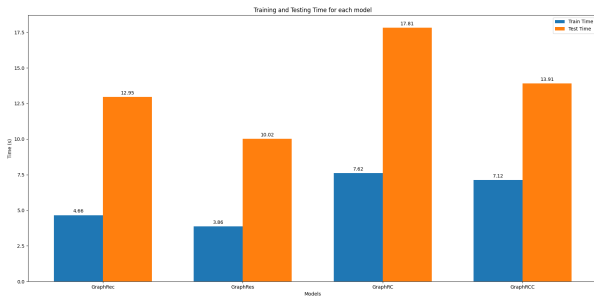


图 12: 相同数据集在四个模型的运行时间比较

① 课题背景

② 基于 GAT 的推荐系统

③ 实验研究

④ 参考文献

- [1] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations*, 2017.
- [2] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.
- [3] W. Fan and et al., "Graph neural networks for social recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 7370–7377.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.
- [5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
- [6] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 173–182.

Thanks!