

# ECNU

ICDM:大规模电商图上的风险商品检测

## Sjkkxcoe团队技术分享

队员：黄逸宏

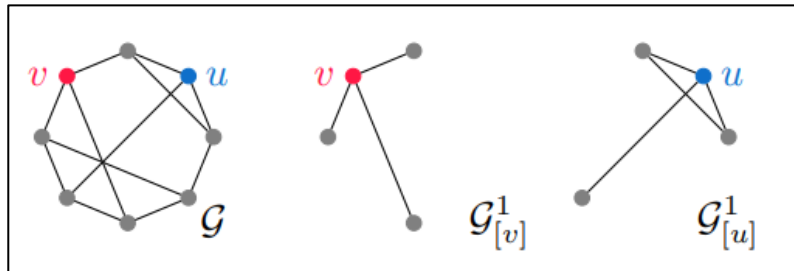
# 目录

- 1. 初赛方案
  - 解耦跳数和层数
  - 优化采样
  - 参数设置
- 2. 复赛方案
  - 利用metapath进行数据增强
- 3. 可能的改进点

# 解耦跳数和层数

01 / 14

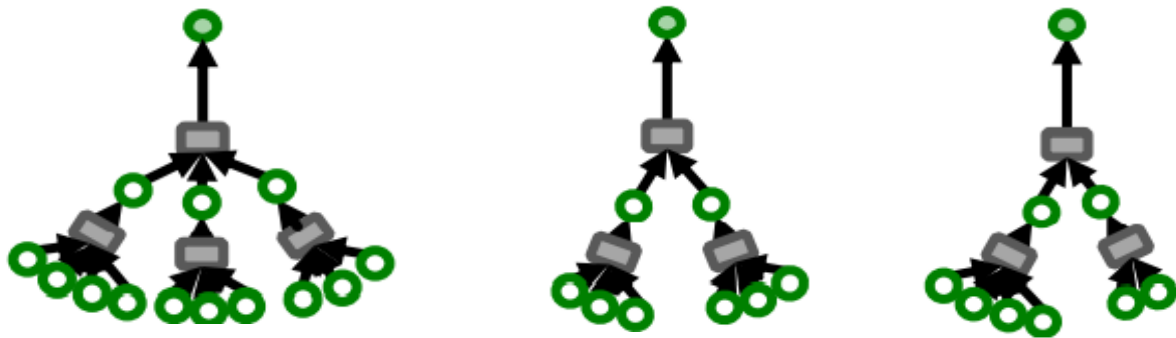
## Decoupling the Depth and Scope of Graph Neural Networks (NeurIPS 2021)



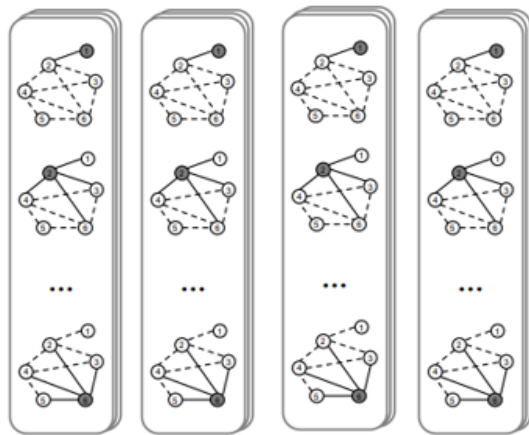
简单地解耦采样跳数和GNN的深度， 可以提高模型的表现。

Pyg 的ShadownKHopSampler完美实现， 但考虑到效率， 使用NeighborLoader进行替代， 效果差不多。

Mini-batch 采样



跳数 = 2



$\hat{y}$

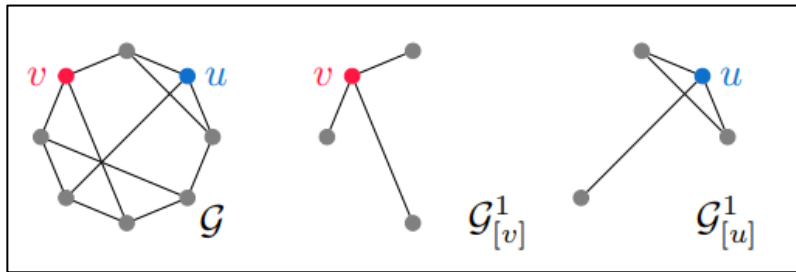
模型深度 = 4

结果:

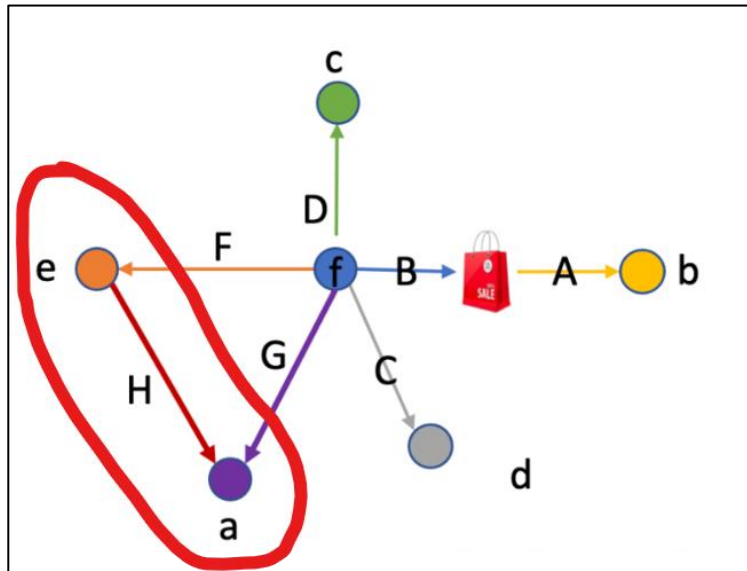
	说明	val_ap
rgcn_sage	baseline	0.95953
rgcn_4	rgcn 增加到 4 层, hops=2	0.96081

# 优化采样

01 / 14



设置采样跳数为2，同时每次将所有mini-batch中两跳节点 $e$ 和节点 $a$ 之间关系 $H$ 的边，全部收录进来。时间效率上几乎没有区别。



结果：

	说明	val_ap
rgcn_sage	baseline	0.95953
rgcn_4	rgcn 增加到 4 层, hops=2	0.96081
ResRGCN_4	rgcn 更换为 resRgcn	0.95975
ResRGCN_4_H	使用方案二加入 H 关系	0.96192

# 初赛及复赛参数设置

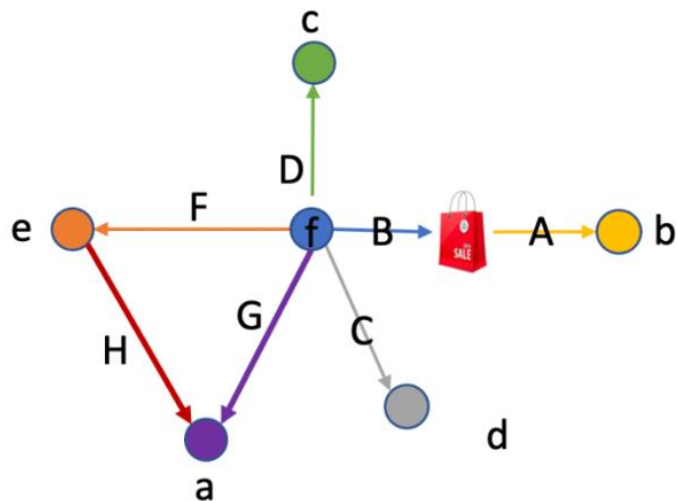
隐藏层维度	256
学习率	0.001
BatchSize	200
模型层数	4
采样的邻居	每种关系采样 150 个，共采样 2 跳
RGCN 的 base	8
库	pyg



# 数据增强- 随机组合metapath作为增强样本

01 / 14

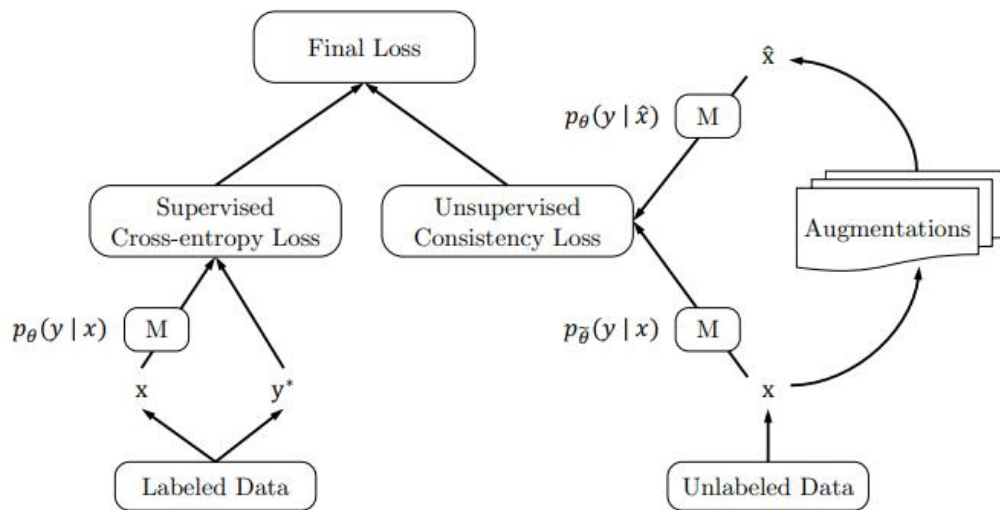
MetaPath 是指一条包含relation序列的路径



Id	MetaPath	MAX_VAL_Score
1	( a – f – item )	0.9471 (0.94714)
2	( c – f – item )	0.9426
3	( f – item )	0.9416
4	( item – f – item )	0.9451
5	( b – item )	0.9346
6	( item – b – item )	0.9380
7	( d – f- item )	0.9418
8	( e – f -item )	0.9439

# 数据增强框架

Unsupervised Data Augmentation for Consistency Training (NeurIPS 2020)



# 损失函数

01 / 14

$$\min_{\theta} J(\theta) = E_{x,y \in L} [-\log p_{\theta}(y|x)] + \lambda E_{x \in U} E_{\hat{x} \sim q(\hat{x}|x)} [D_{KL}(p_{\tilde{\theta}}(y|x) || p_{\theta}(y|\hat{x}))]$$

$\tilde{\theta}$ 代表不对这里的参数求导

## 额外的训练技巧

$$\frac{1}{|B|} \sum_{x \in B} \underbrace{I(\max_{y'} p_{\tilde{\theta}}(y' | x) > \beta)}_{\text{Confidence-based masking}} \text{CE} \left( p_{\tilde{\theta}}^{(sharp)}(y | x) || p_{\theta}(y | \hat{x}) \right)$$

$$\underbrace{p_{\tilde{\theta}}^{(sharp)}(y | x) = \frac{\exp(z_y/\tau)}{\sum_{y'} \exp(z_{y'}/\tau)}}_{\text{Sharpening Predictions}}$$

Sharpening Predictions

结果:

	说明	val_ap
rgcn_sage	baseline	0.95953
rgcn_4	rgcn 增加到 4 层, hops=2	0.96081
ResRGCN_4	rgcn 更换为 resRgcN	0.95975
ResRGCN_4_H 0.914016	使用方案二加入 H 关系	0.96192
ResRGCN_4_H_meta 0.917587	加入随机组合 metapath 的数据增强	0.96571

# 可能的改进点

01 / 14

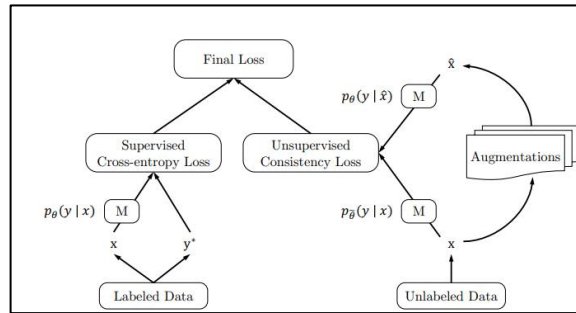
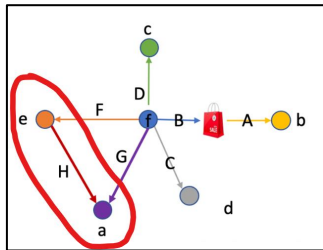
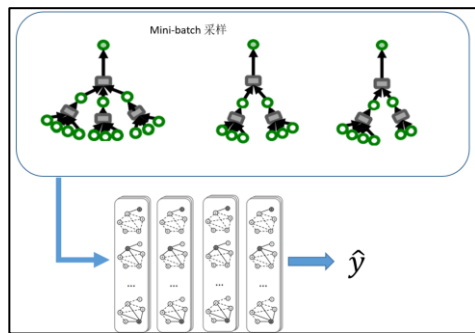
## 1. 没有利用上无标签数据

先前的数据增强的方法可以利用上。

## 2. 没有对节点的属性进行补全

**Heterogeneous Graph Neural Network via Attribute Completion**, (WWW 2021 best paper)

# 总结



结果:

	说明	val_ap
rgcn_sage	baseline	0.95953
rgcn_4	rgcn 增加到 4 层, hops=2	0.96081
ResRGCN_4	rgcn 更换为 resRgcN	0.95975
ResRGCN_4_H	使用方案二加入 H 关系	0.96192
ResRGCN_4_H_meta	加入随机组合 metapath 的数据增强	0.96571

# 感谢聆听！ Q & A