



# 数字媒体技术基础

**Meng Yang**

[www.smartllv.com](http://www.smartllv.com)

**SUN YAT-SEN University**



**机器智能与先进计算教  
育部重点实验室**



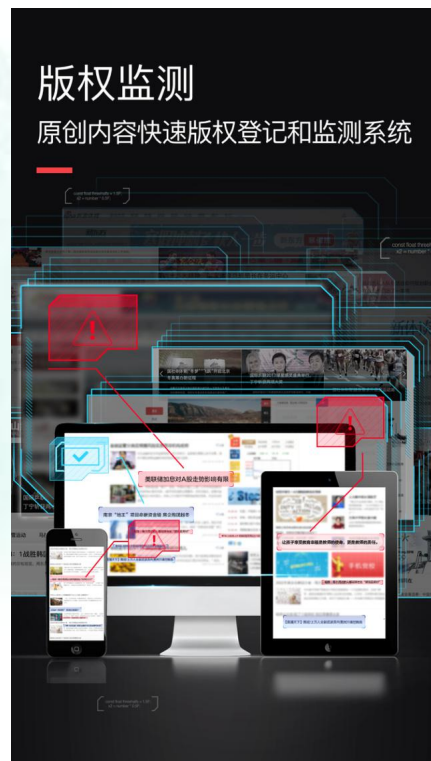
**智能视觉语言  
学习研究组**

## 第十三章 智能新媒体处理技术

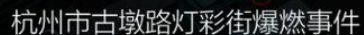
# 第十三章 智能新媒体处理技术

## 新闻采集能力显著提升

新华社开发的“媒体大脑”，能实现对上亿个网页进行快速扫描，并对网页中的文本、图像、视频等信息进行采集处理，而且能对信息是否与两会相关进行准确判断，甚至能够判定哪些议题可以成为热点新闻。

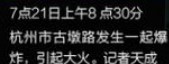
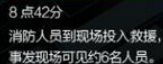
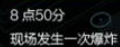


# 智能媒体生产平台



⌚ 2017年7月21日8时30分

📍 杭州市古墩路灯彩街交叉口



## 结合新闻大数据的智能分发平台





# 第十三章 智能新媒体处理技术



## □ 内容生产更加智能化

写作机器人被广泛应用，比如腾讯的Dreamwriter、今日头条的Xiaomingbot、新华社的快笔小新等。

### “快笔小新”：新华社第一位机器人记者

2017-10-09 10:41

2015年11月7日，在新华社84岁生日之际，新华社的编辑记者们迎来了一位特别的同事，他的名字叫“快笔小新”。这是新华社第一位机器人记者。经过新华社技术部门一年多的不断迭代开发，快笔小新具备了更强大的“写稿”能力。

快笔小新上线运行后，7×24小时不间断工作。每逢重要体育赛事结束，或股市开收盘、突破整数关口，或是上市公司财报季报等金融信息发布之时，就会自动根据所公布的信息快速生成新闻稿件，编辑记者在编辑系统对快笔小新生成的稿件进行核对后正式签发。



[https://www.sohu.com/a/196905347\\_644338](https://www.sohu.com/a/196905347_644338)

# 第十三章 智能新媒体处理技术



## 信息分发能力大幅提升

新华社的“媒体大脑”能够实现“用户画像”功能，掌握用户的位置变化及行为习惯，进而对信息内容进行准确分发。



# 第十三章 智能新媒体处理技术



## 智能新媒体时代，需要更强大的媒体处理技术

采集途径多样化

采集能力提升

媒体形式多样化

分发途径多样化

分发能力提升

数据量更大，需要提升  
分类能力、检索能力、  
推荐能力



- ❑ 13.1 智能新媒体分类技术
- ❑ 13.2 智能新媒体检索技术
- ❑ 3.3 智能新媒体推荐技术



## 13.1 智能新媒体分类技术

# 为什么需要对新媒体分类？

**Jerry**很关心足球新闻，可当他打开一个新闻网站，却发现这里的新闻眼花缭乱，一下子根本找不到关于足球的新闻……



- 新闻自动分类



- 情感分类



- 谣言识别



今天是马云的生日，复制这条信息 WmKukX765S到支付宝，获得红包后并将信息转发给10个微信好友，您的支付宝账号钻石等级将被永久点亮。

- 不良信息识别



- .....

- 如何建立媒体之间的交互
- 缺乏有监督数据
- .....



# 新媒体分类主流方法概述

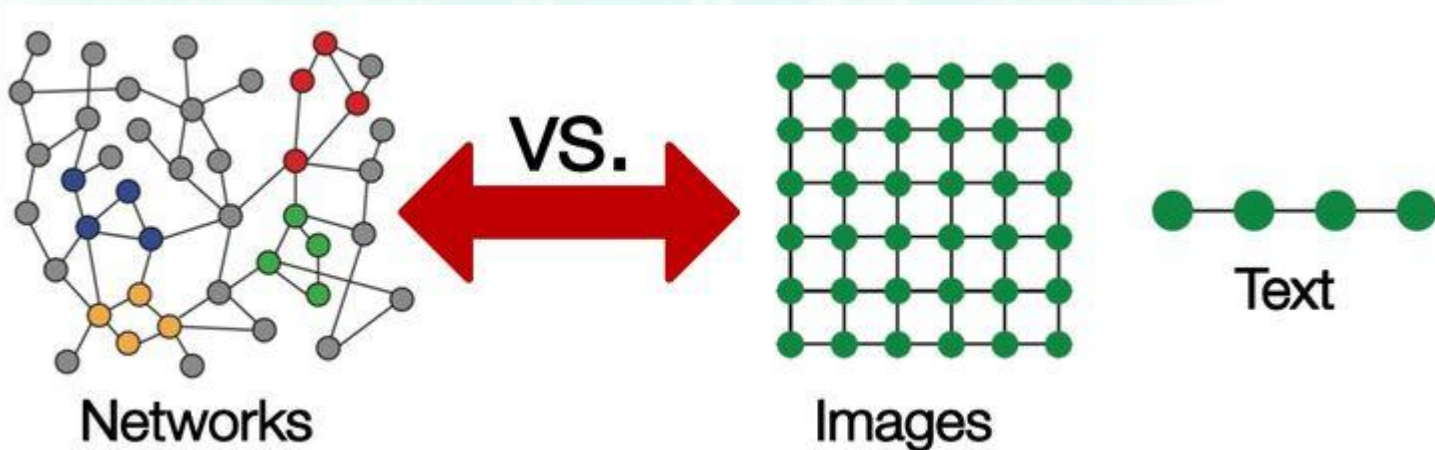


- 传统机器学习模型：随机森林，SVM，KNN等；
- 深度学习模型：将文本中的单词编码，得到文本的整体表示，然后进行分类，如fastText、TextCNN等；
- 图神经网络方法：通常以文本为节点，建立图神经网络，以加强文本之间的交互。



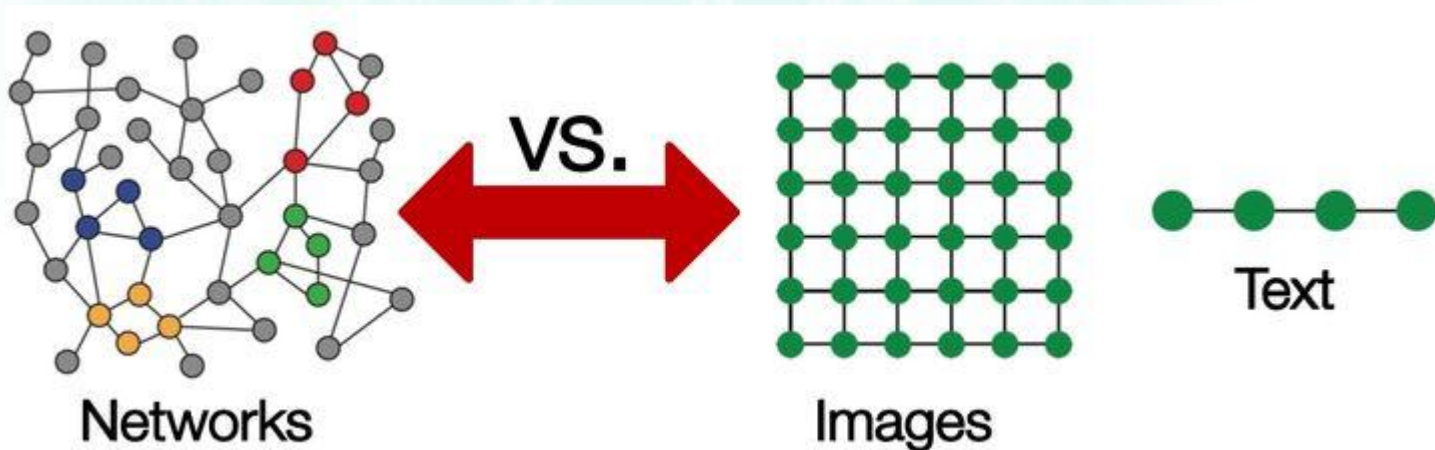
# 图神经网络 (GNN) 介绍

- 对于单个图片/文本，通常以向量/矩阵等来表示。例如，每个数值表示一个像素。
- 而对于海量的智能新媒体，我们还希望构建多条媒体之间的联系。一个很直接的方法就是引入图结构。



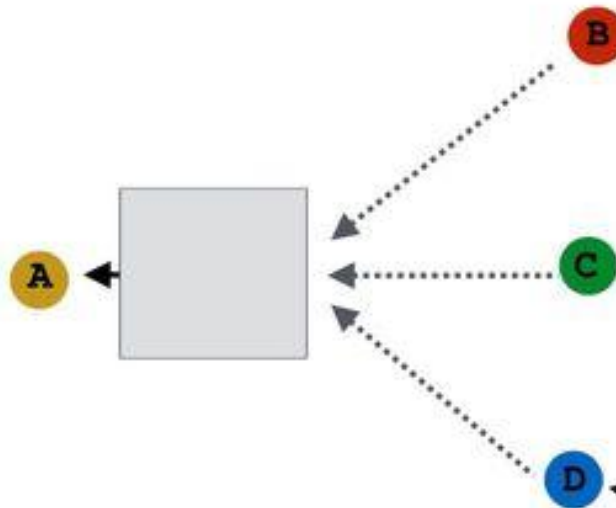
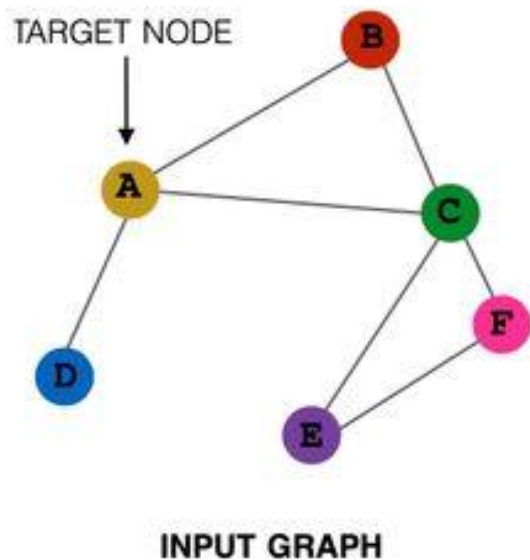
# 图神经网络（GNN）介绍

- 在图结构中，每个节点（可代表一条新闻）通过连边和邻居节点相连，节点之间的信息只能通过连边来传播。
- 相比起直接将所有节点相连（全连通图），利用外部信息来构建GNN更加合理。



# 图神经网络（GNN）介绍

- GNN的信息传递策略：对于节点A，其邻居节点为B、C、D，那么节点A就可被B、C、D更新。

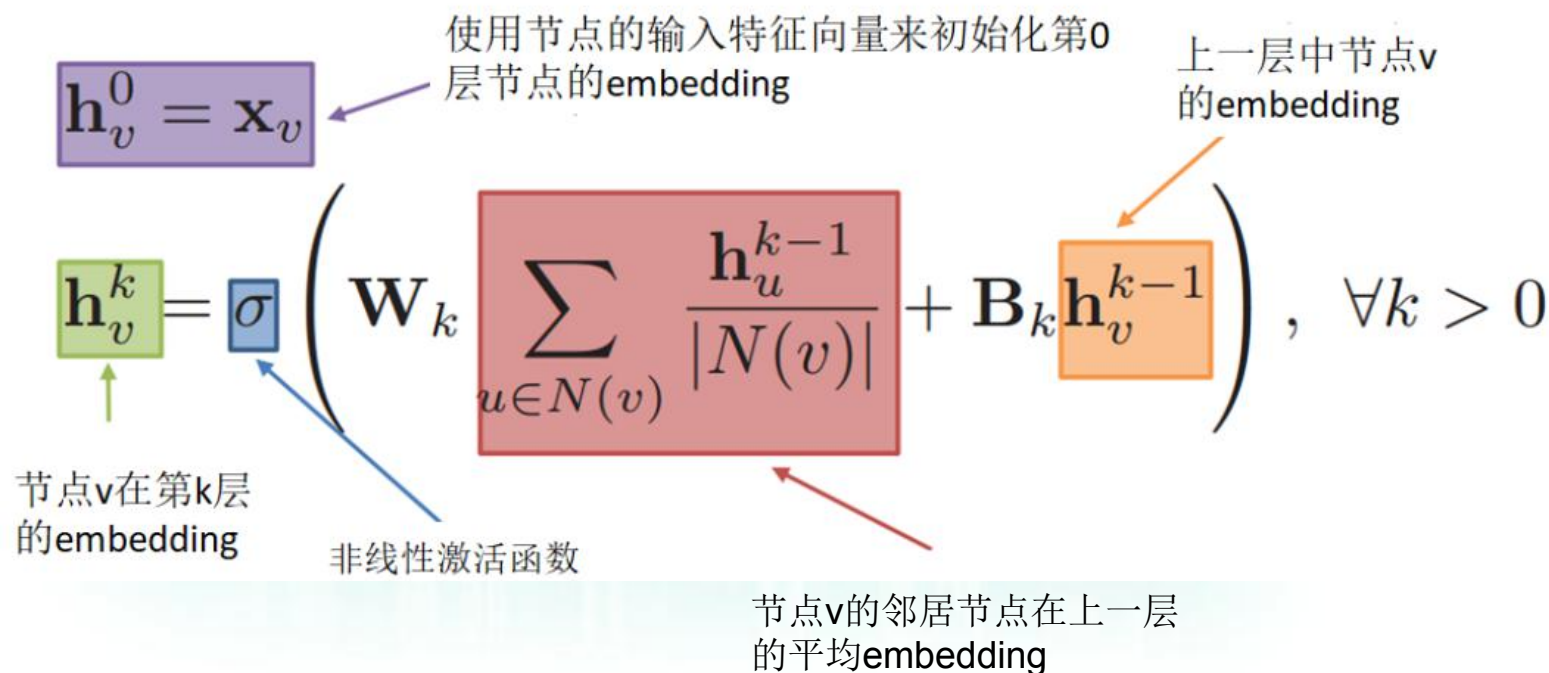


# 图神经网络（GNN）介绍



问题？

- 对所有节点都执行类似的操作，就能让节点都包含邻居的信息。

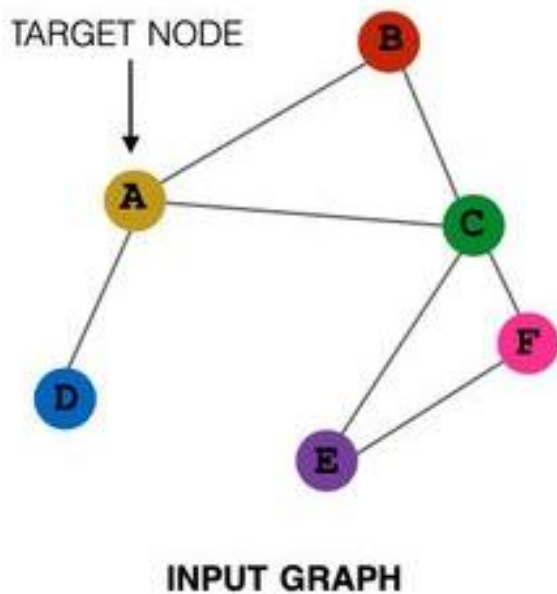


具体的更新函数会因为具体的模型有所不同，经典模型包括图卷积网络GCN、图注意力网络GAT、GraphSAGE等。

# 图神经网络（GNN）介绍



问题：对节点A来说，需要多少次迭代更新，才能接收到节点E的信息？



2次。第一次更新能让E的信息传递到C和F，第二次更新能让C信息传递到A、B、E和F。

问题？



## 基于GNN的文本分类——TextGCN

- 节点类型:

- ① 文档节点（语料库中的所有文档）
- ② 单词节点

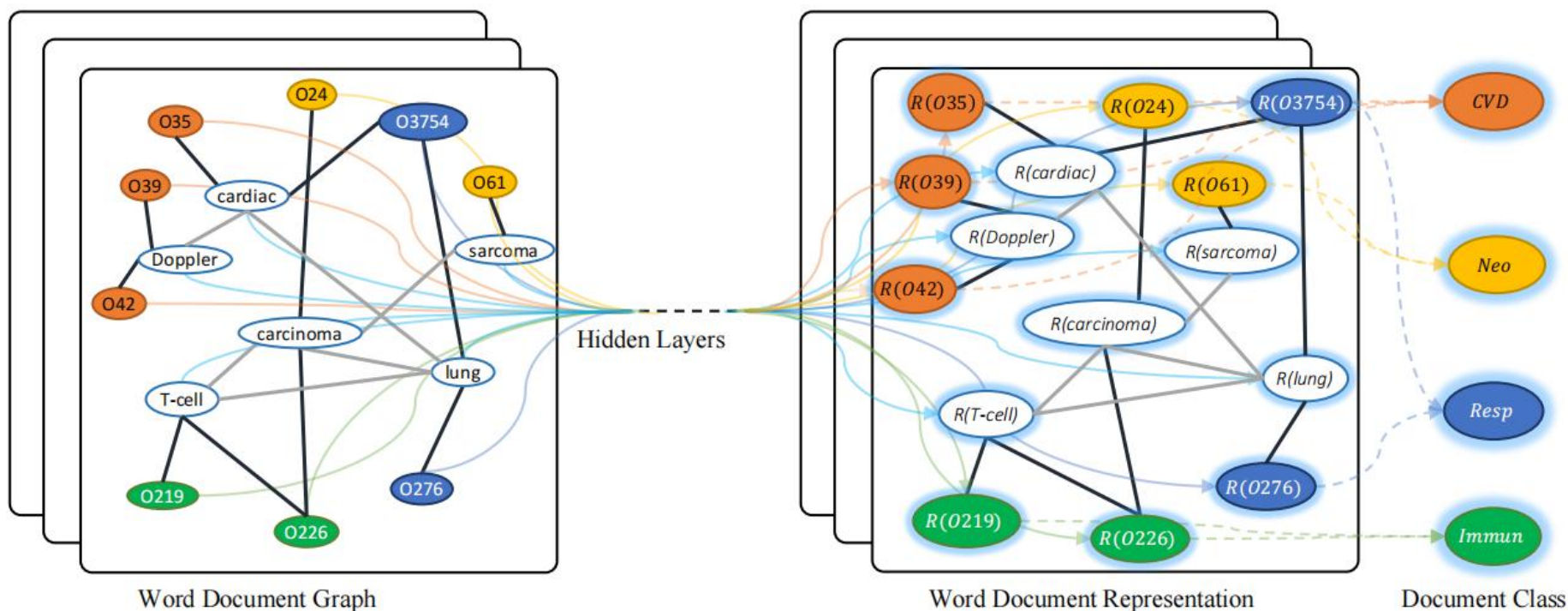
- 连边构建（邻接矩阵）:

- ① 文档-单词连边：用TF-IDF值
- ② 单词-单词连边：用PMI（point-wise mutual information）值

$$A_{ij} = \begin{cases} \text{PMI}(i, j) & i, j \text{ are words, } \text{PMI}(i, j) > 0 \\ \text{TF-IDF}_{ij} & i \text{ is document, } j \text{ is word} \\ 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

Yao L, Mao C, Luo Y. Graph convolutional networks for text classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 7370-7377.

## 基于GNN的文本分类——TextGCN



Yao L, Mao C, Luo Y. Graph convolutional networks for text classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 7370-7377.

## 基于GNN的文本分类——TextGCN

第二层GCN

第一层GCN

$$Z = \text{softmax}(\tilde{A} \text{ReLU}(\tilde{A}XW_0)W_1)$$

邻接矩阵的正则化表示

$X$ : 每个节点的特征

对所有文档 ( $\mathcal{Y}_D$ ) 计算交叉熵损失函数:

$$\mathcal{L} = - \sum_{d \in \mathcal{Y}_D} Y_d \ln Z_d$$

ground-truth

## TextGCN存在的问题

- 为整个语料库构建一个异构图，图十分庞大且无法在线添加新节点，即训练完成就很难再修改了。
- 连边的权重是固定的，限制了连边的表达能力。

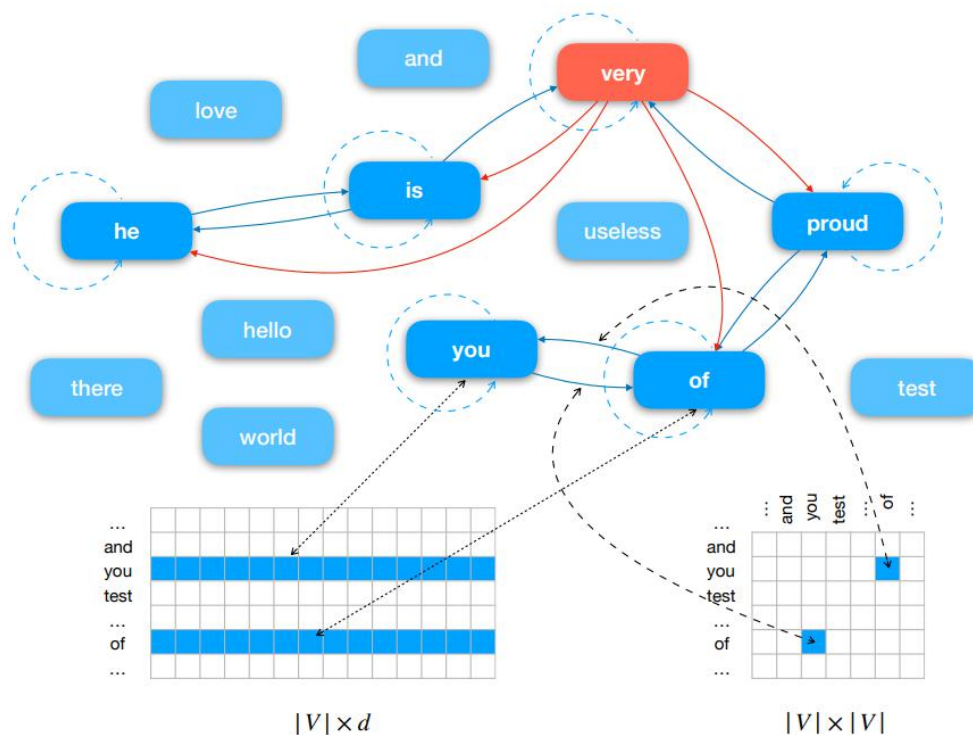
## 基于GNN的文本分类——TextLevelGNN

- 为每个文本单独构建一个图，文本中的单词作为节点（每个单词与其上下文的 $p$ 个词相连）；
- 节点（单词）表示以及连边权重（单词之间的权重）被整个语料库共享；
- 由于无需对整个文本编码，因此，每进来一篇新的文本，都能输入到模型中训练或测试。



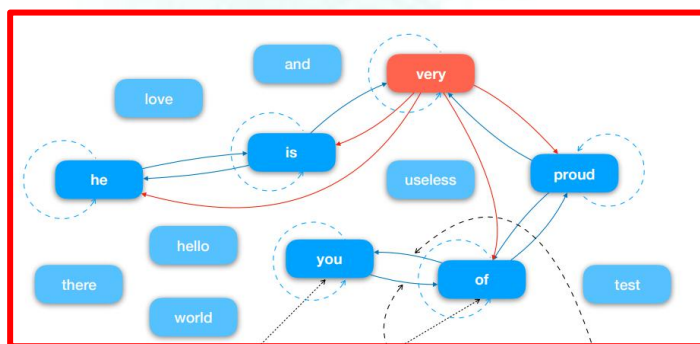
## 基于GNN的文本分类——TextLevelGNN

- 以文本“he is very proud of you”为例，“very”的窗口大小 $p=2$  (标红)，其他词的窗口大小 $p=1$ ；
- 下方左侧：节点表示矩阵，整个语料库共享；
- 下方右侧：单词之间的权重矩阵，整个语料库共享。



## 基于GNN的文本分类——TextLevelGNN

- 如何将GNN转化为文本表示？



文本表示

$$y_i = \text{softmax}(\text{Relu}(\mathbf{W} \sum_{n \in N_i} \mathbf{r}'_n + \mathbf{b}))$$

对所有节点(单词) 的表示求和

Huang L, Ma D, Li S, et al. Text level graph neural network for text classification[J]. arXiv preprint arXiv:1910.02356, 2019.

## 基于GNN的文本分类——TextLevelGNN

- 训练时，同样采用交叉熵损失函数：

$$\text{loss} = -g_i \log y_i$$

ground-truth

模型预测的分类概率

## 13.2 智能新媒体检索技术

# 为什么需要新媒体检索?

Tom注册了微博，他想看看是谁在微博上说他的坏话、发他的丑照。他在整个微博网站中一条一条地找，找到昏睡过去还没找到……





- 淘宝商品搜索



- 公众号搜索

- 微博搜索

- ■■■■■■



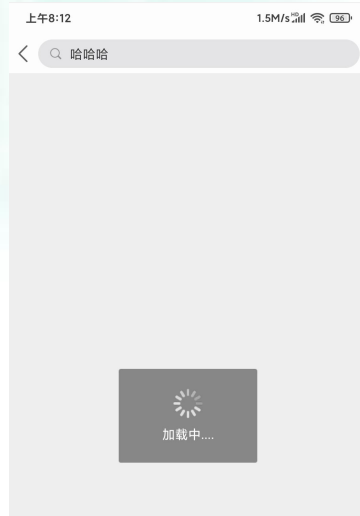
- 海量数据需要被检索



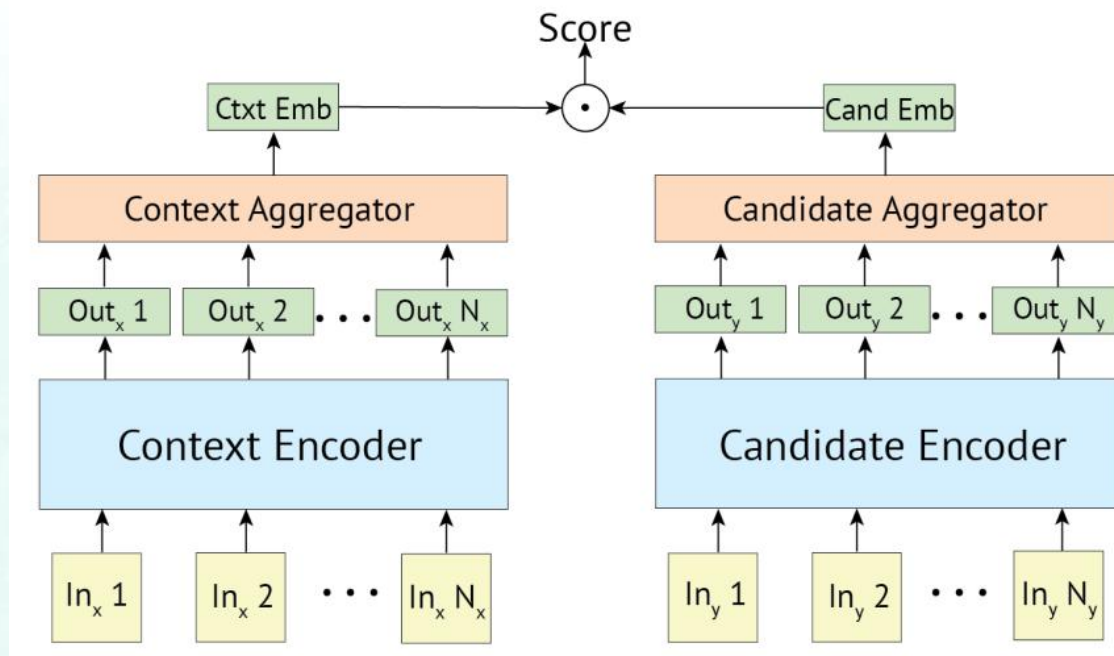
截至2018年12月，新浪微博全站发布博文超过2000亿条、图片500亿张、视频4亿个、评论和赞总量近5000亿。

- 检索效率：太慢会降低用户体验

- .....



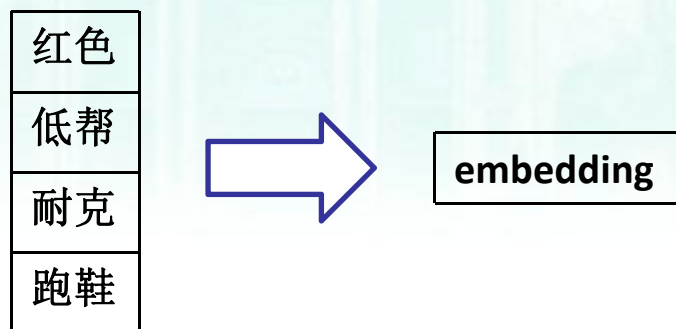
## Bi-Encoder:



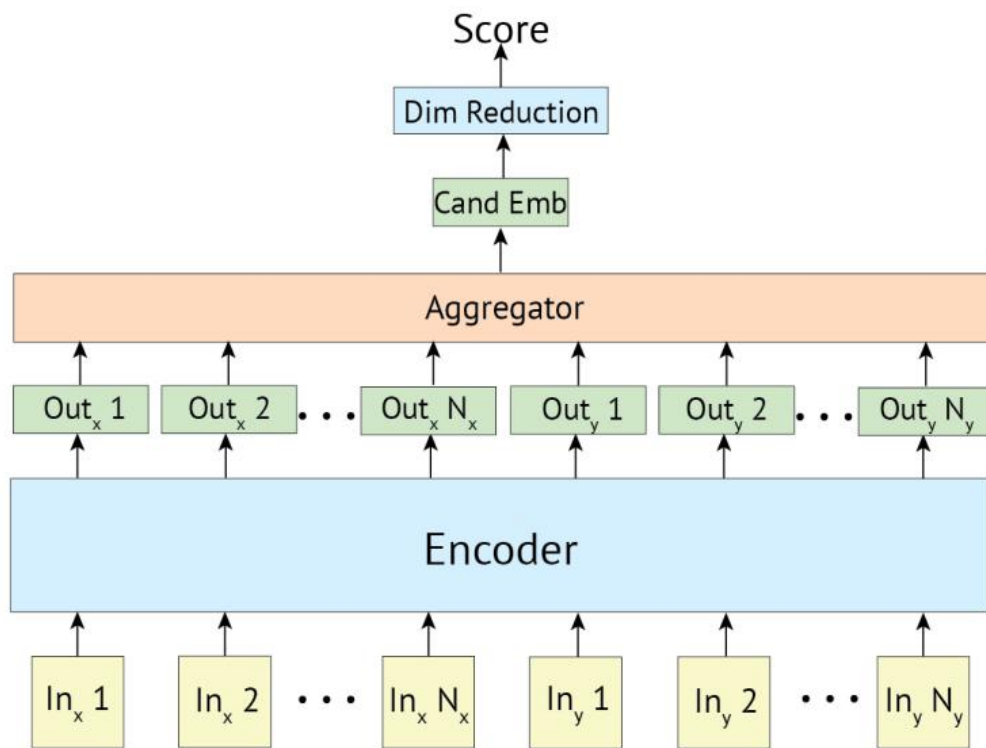
- 左侧对query编码（例如在搜索框输入的关键词）；
- 右侧对condidate编码（例如候选文章、候选图片）；
- 最终，根据点积或cosine计算相似度。

## Bi-Encoder存在的问题:

- **query和candidate之间缺乏深度交互!**
  - 例如, 在淘宝搜索栏输入query “红色低帮耐克跑鞋”, Encoder首先会将“红色”、“低帮”、“耐克”、“跑鞋”对应编码成共4个词向量, 然后将它们压缩成一个向量表示。这时候再和candidate商品匹配, 会丢失一些信息。
  - 另一个角度来说, 有点像word2vec静态词向量: 即使一个词有多种语义, 它的所有语义也不得不塞进一个固定的词向量。



## Cross-Encoder:



- 将query和candidate直接拼接，输入模型中编码，计算得分
- （类似于Video-BERT）

问题：Cross-Encoder存在什么问题？

问题？

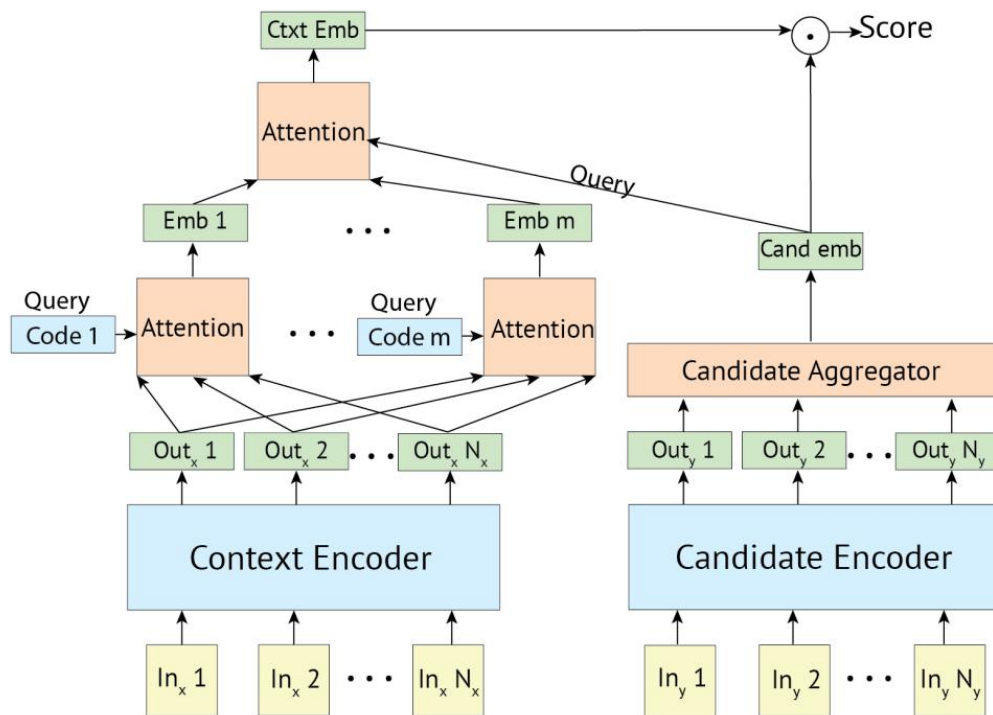
预测时复杂度过高，每来一个candidate，都需要对query重新编码！

例如，如果有N个candidates，则在Encoder模块需要对query+candidate编码N次。





## Poly-Encoder:



- 相比Bi-Encoder将query压缩成单个向量，Poly-Encoder将query编码成  $m$  ( $m$ 是超参) 个向量，类似于multi-head机制。

## Poly-Encoder:

- 这是Bi-Encoder和Cross-Encoder的折中：相比Bi-Encoder，交互更深；相比Cross-encoder，检索效率更高。
- 将query映射到多个子空间，可理解为在不同的维度保留有用信息。例如，在淘宝搜索，不同的维度可能分别侧重于商品类型、颜色、形状、尺寸等。

## 13.3 智能新媒体推荐技术

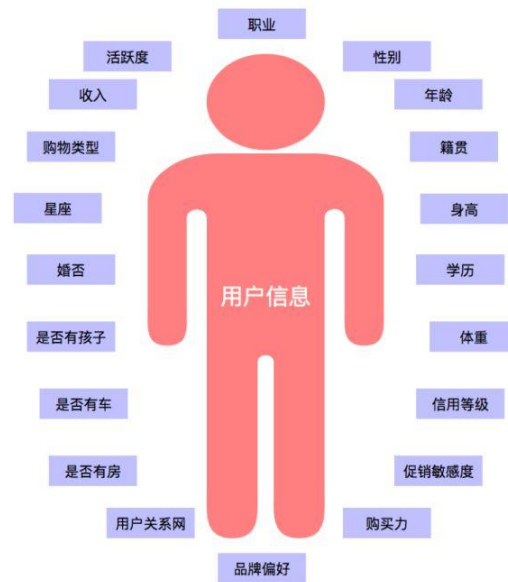
# 为什么需要新媒体推荐技术？

- Jerry今天有点无聊，于是它打开了淘宝，首页显示的却是一个捕鼠夹……它继续往下滑，依次看到了猫粮、猫的衣服、猫的日用品……



- 淘宝商品个性化推荐
- 抖音短视频短性化推荐
- 新闻个性化定制
- .....

- 海量数据
- 增量处理：微博/新闻每天都更新，如何更好地在模型中表示它们？
- 冷启动问题：一开始，缺少大量用户的点积信息，如何启动推荐系统？
- 个性化推荐、用户画像
- .....





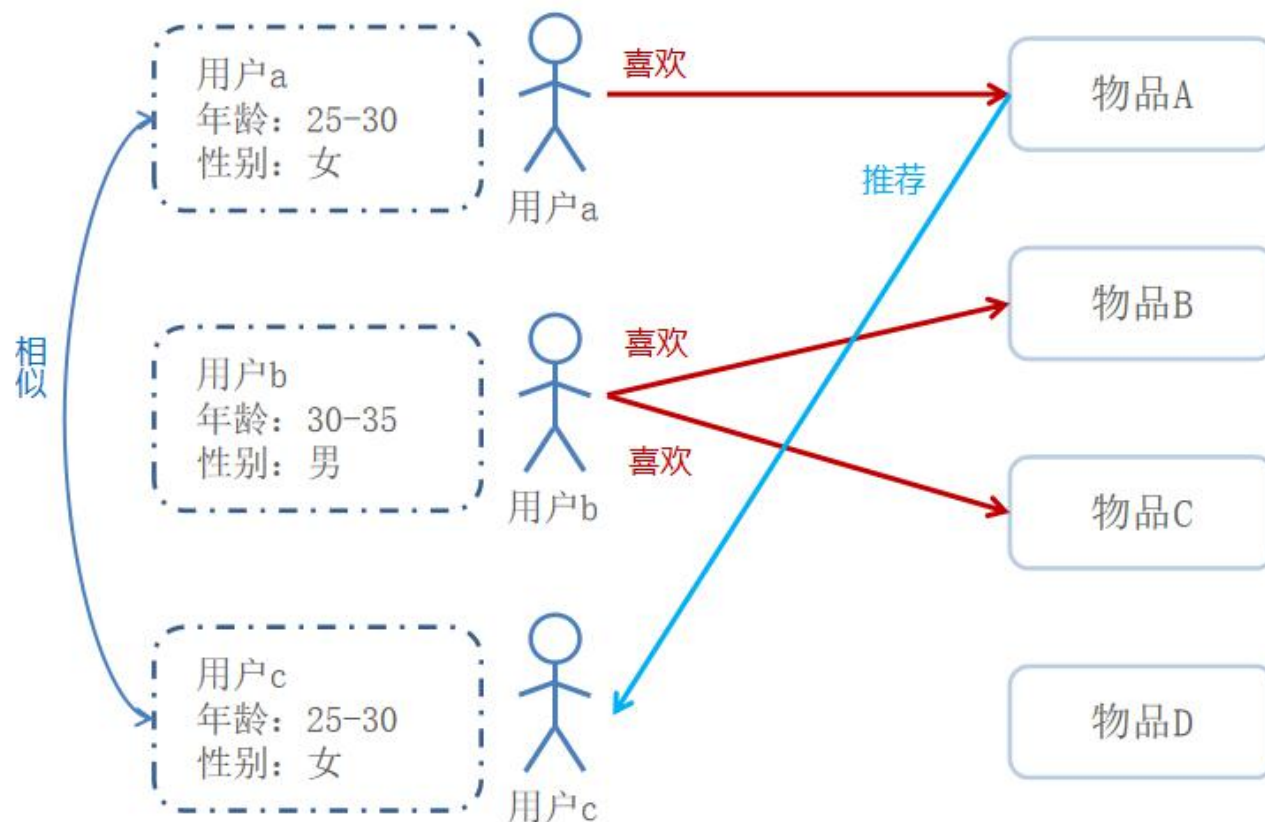
面对海量的媒体信息，从中快速推荐出符合用户特点的物品。

- 从用户的角度来说：解决一些人的“选择恐惧症”，面向没有明确需求的人；
- 从商家的角度来说：能更有针对性地投放广告，提升商品销量或图文阅读量。

# 推荐系统介绍



## 基于用户信息的推荐：



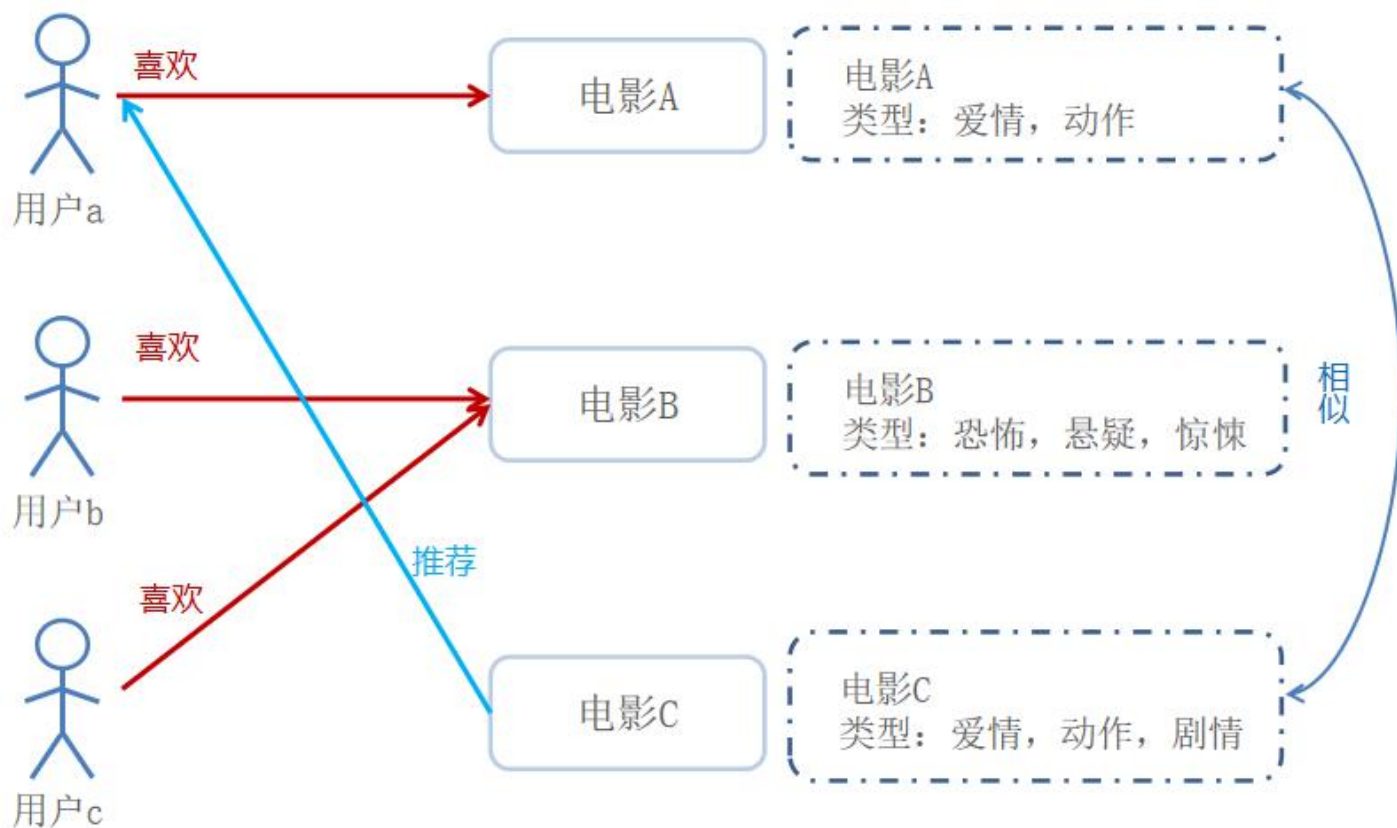
问题：如果你是淘宝APP的产品经理，在用户注册账号时，你觉得收集哪些用户信息对个性化商品推荐有用？为什么？（不考虑用户隐私）

- 性别：男生喜欢游戏机，女生喜欢化妆品
- 年龄：小孩子喜欢玩具，老人喜欢保健品
- 城市：南方会买抽湿机，北方会买羽绒服
- 学历：高学历可能会买专业书籍、科研设备
- 收入：为高收入者推荐贵的东西
- .....

# 推荐系统介绍



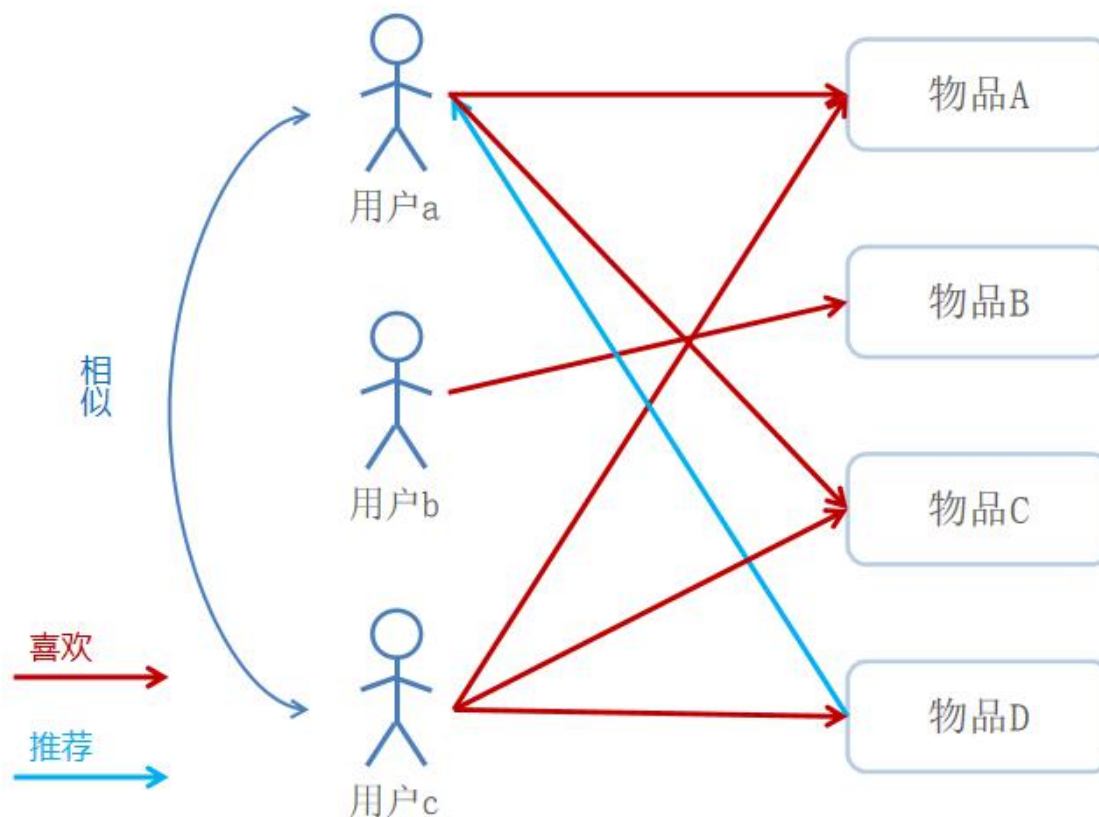
## 基于内容的推荐:



# 推荐系统介绍



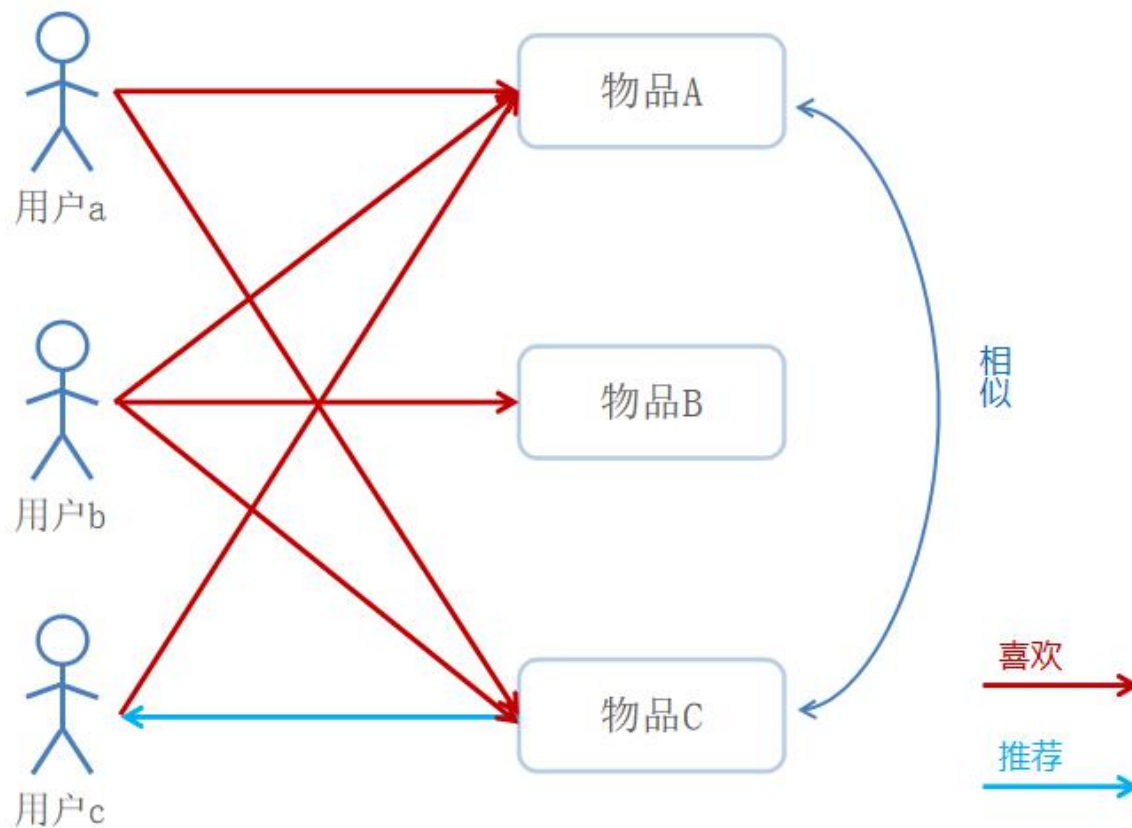
## 基于(用户)协同过滤的推荐:



# 推荐系统介绍



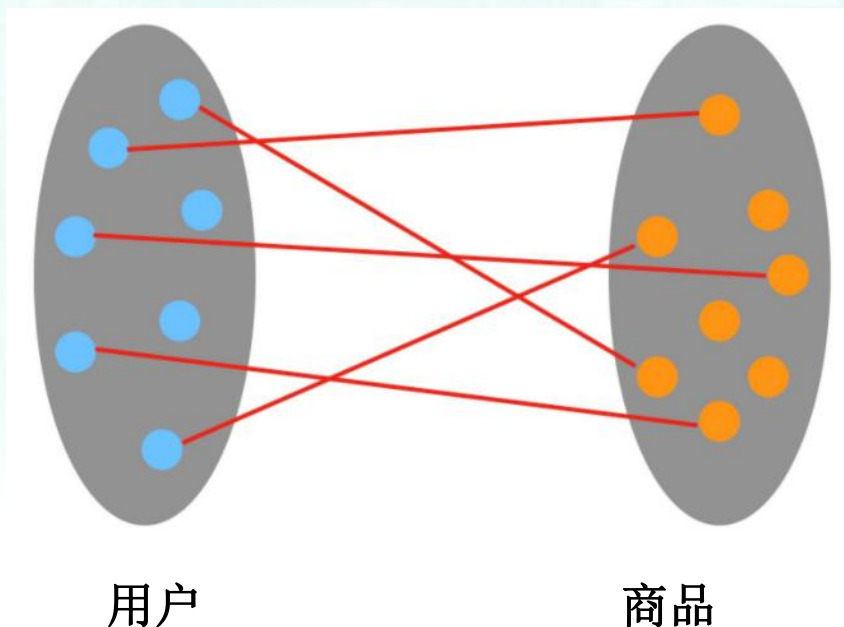
基于(内容/物品)协同过滤的推荐:



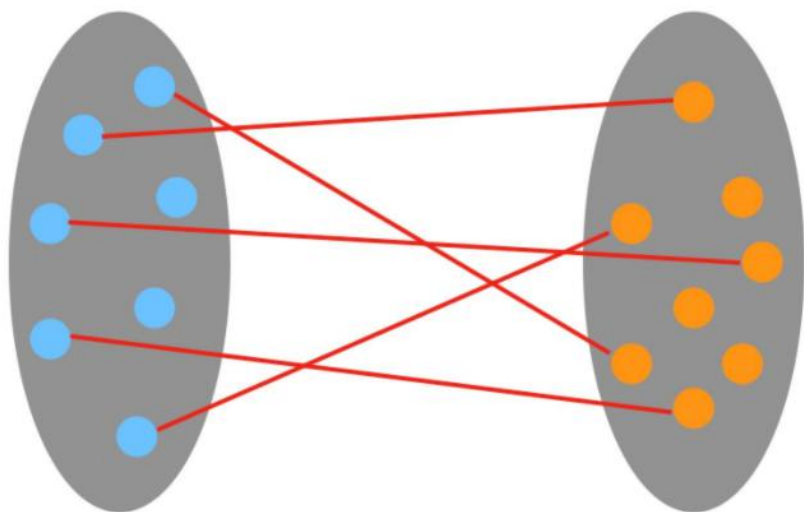


**PinSAGE:** GCN在商业推荐系统首次成功应用，包含数十亿个节点，上百亿条连边。

- 采用二部图结构，即包含两种类型的节点，分别表示用户和商品（也可以是图片、音乐等）。



问题：对于这两类节点，你觉得可以如何定义连边呢？即什么情况下才连接用户 $i$ 和商品 $j$ ？



用户

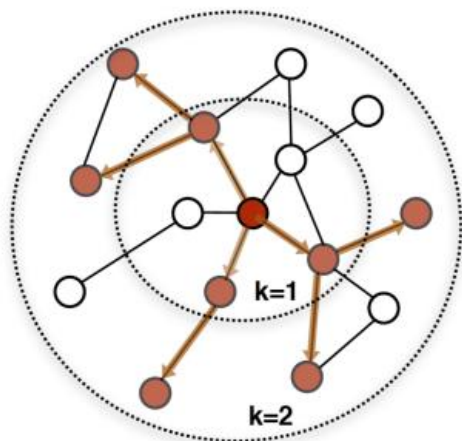
淘宝商品

答案不唯一，例如：

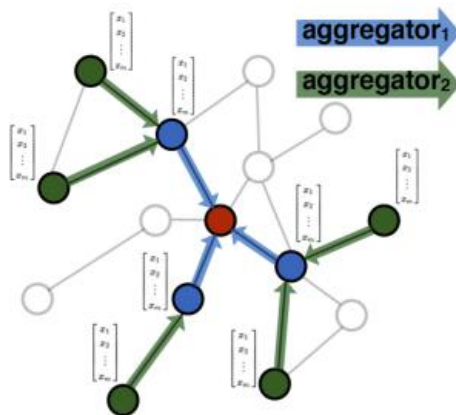
- 若用户 $i$ 点击了商品 $j$ ，则连边；
- 若用户 $i$ 购买了商品 $j$ ，则连边；
- 若用户 $i$ 浏览商品 $j$ 超过30秒，则连边；
- .....

## PinSAGE

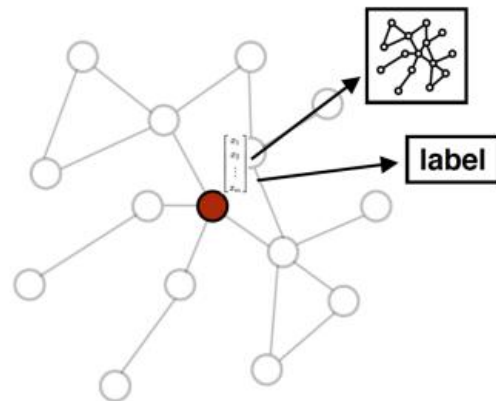
### 训练过程:



(a) 随机游走采样邻居节点



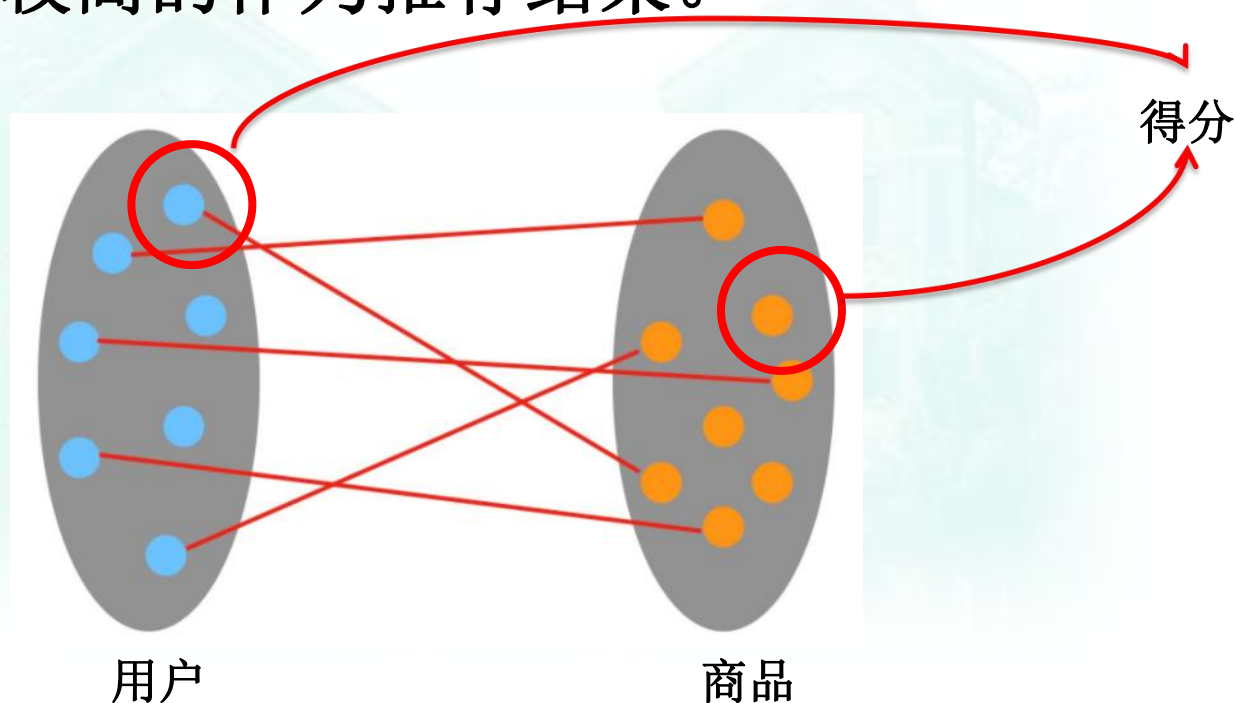
(b) 从邻居节点聚合信息，用于更新自己节点



(c) 有监督地预测自己所属类别，或是无监督地预测图结构（周围哪些节点是自己的邻居）

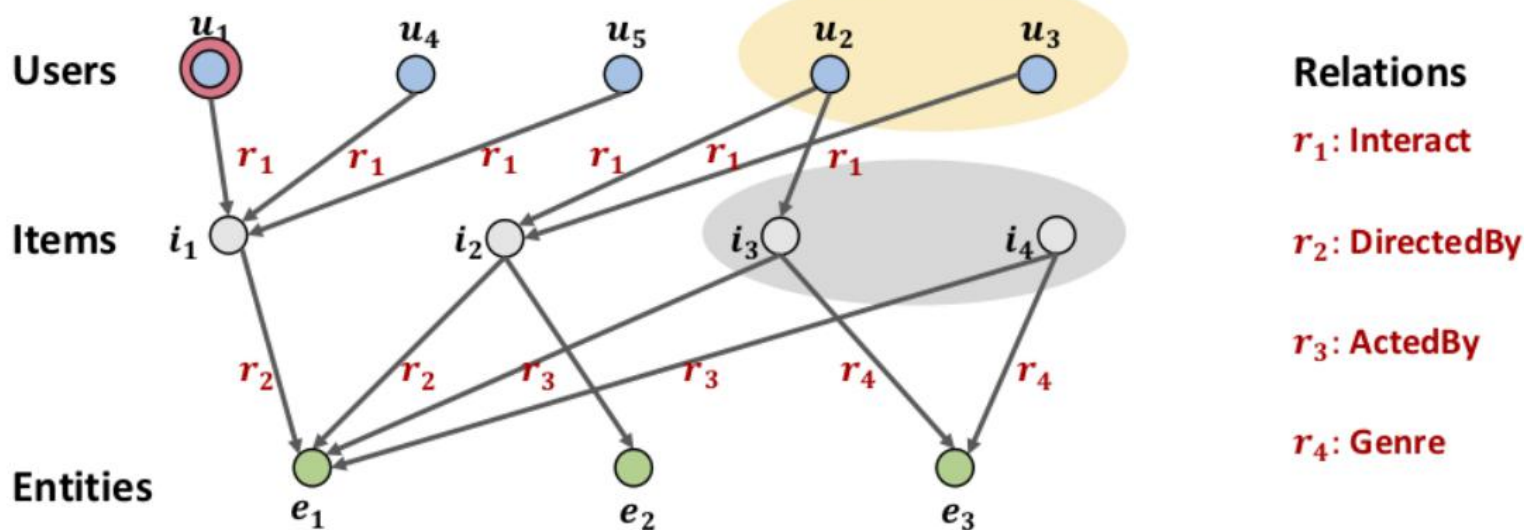
## PinSAGE

预测过程：根据所选用户，计算候选商品的得分，选取较高的作为推荐结果。



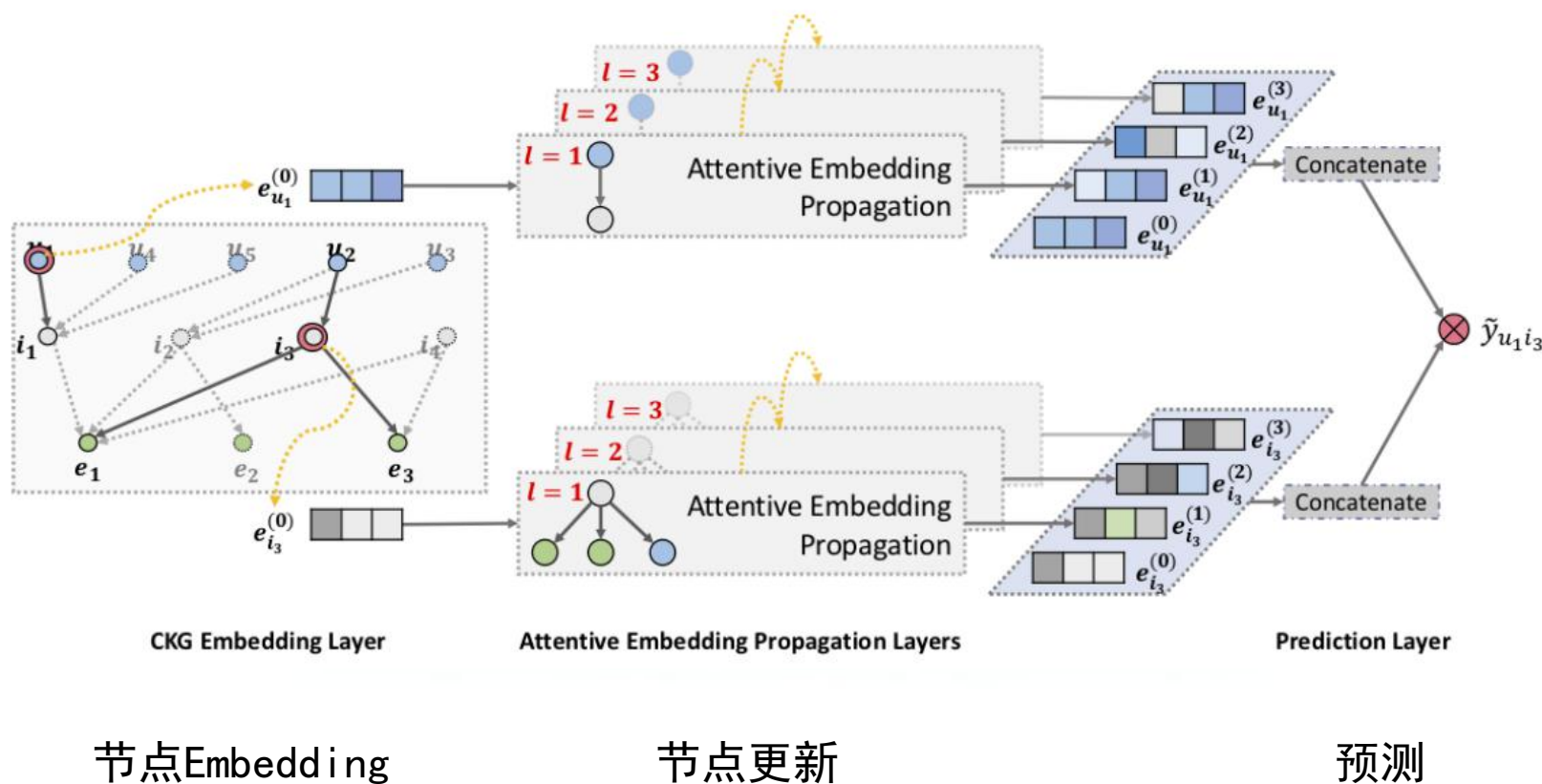
## KGAT——引入外部知识图谱

- 解决冷启动和数据稀疏的问题



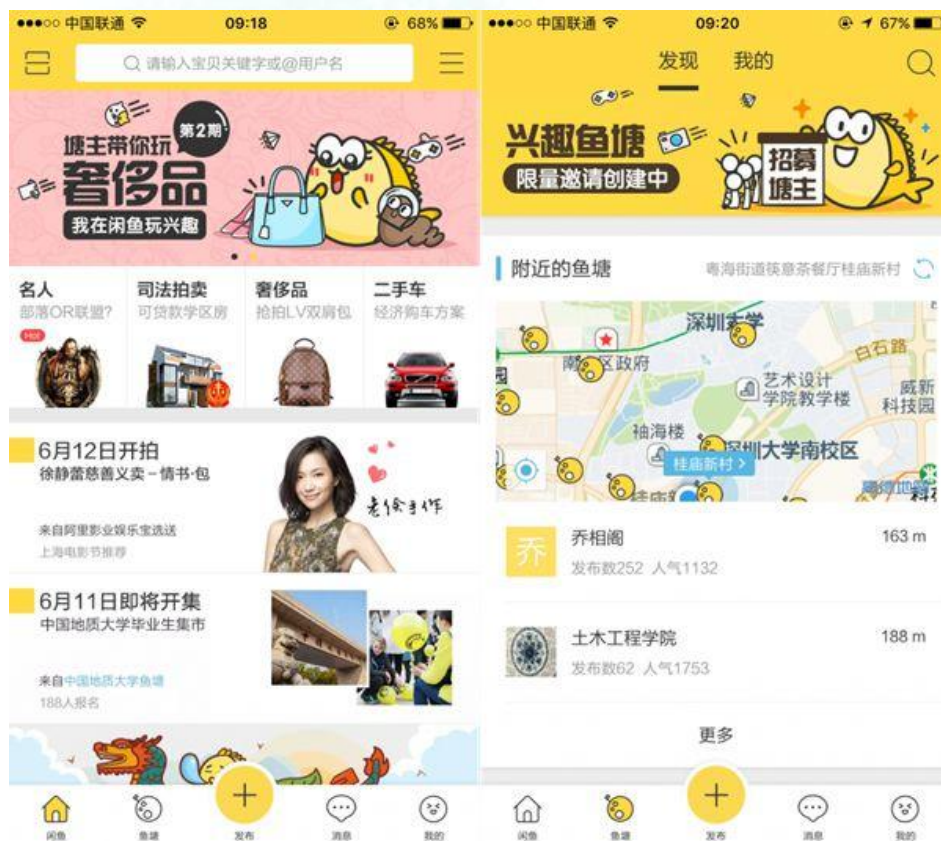
User和Item(例如图片、商品)是经典的二部图，而外部知识图谱是Items和Entities构成的图。本模型在于将这两个图融合为一个图。

## KGAT——引入外部知识图谱



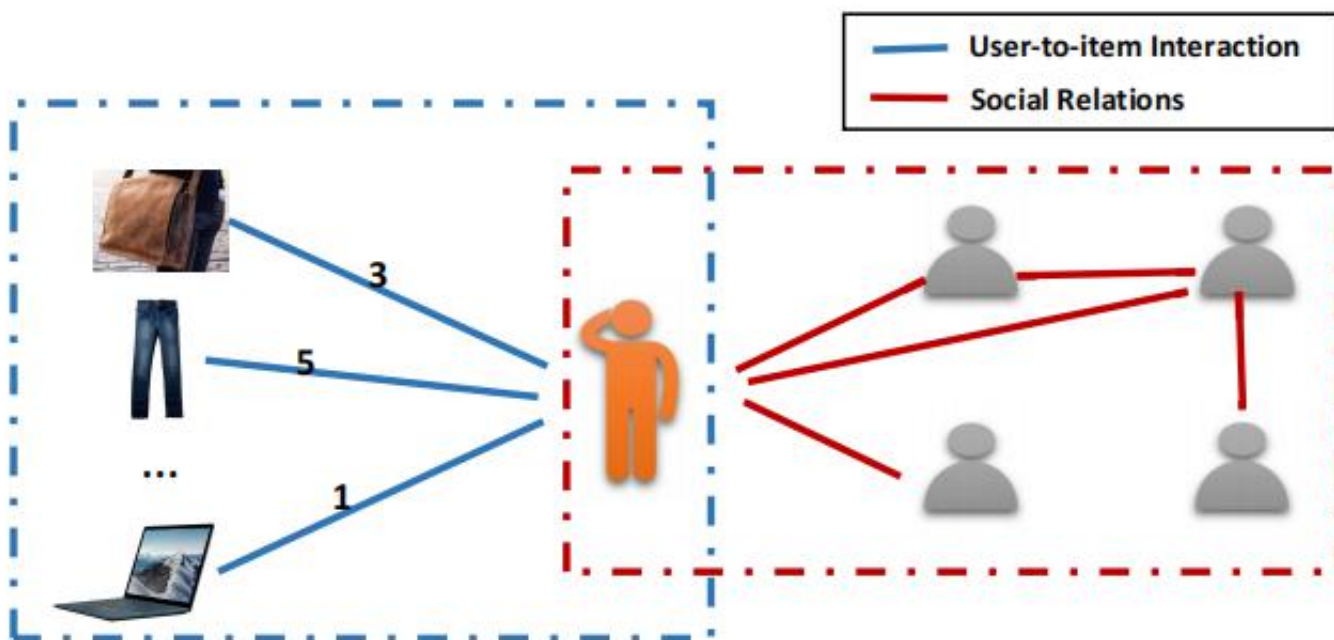


## GraphRec——考虑社交属性



例如，“闲鱼APP”推出的“鱼塘”（多个用户处于相同的社群），那么这些用户之间的交流信息就能被利用上。同一个鱼塘的用户可能会具有相似的喜好。

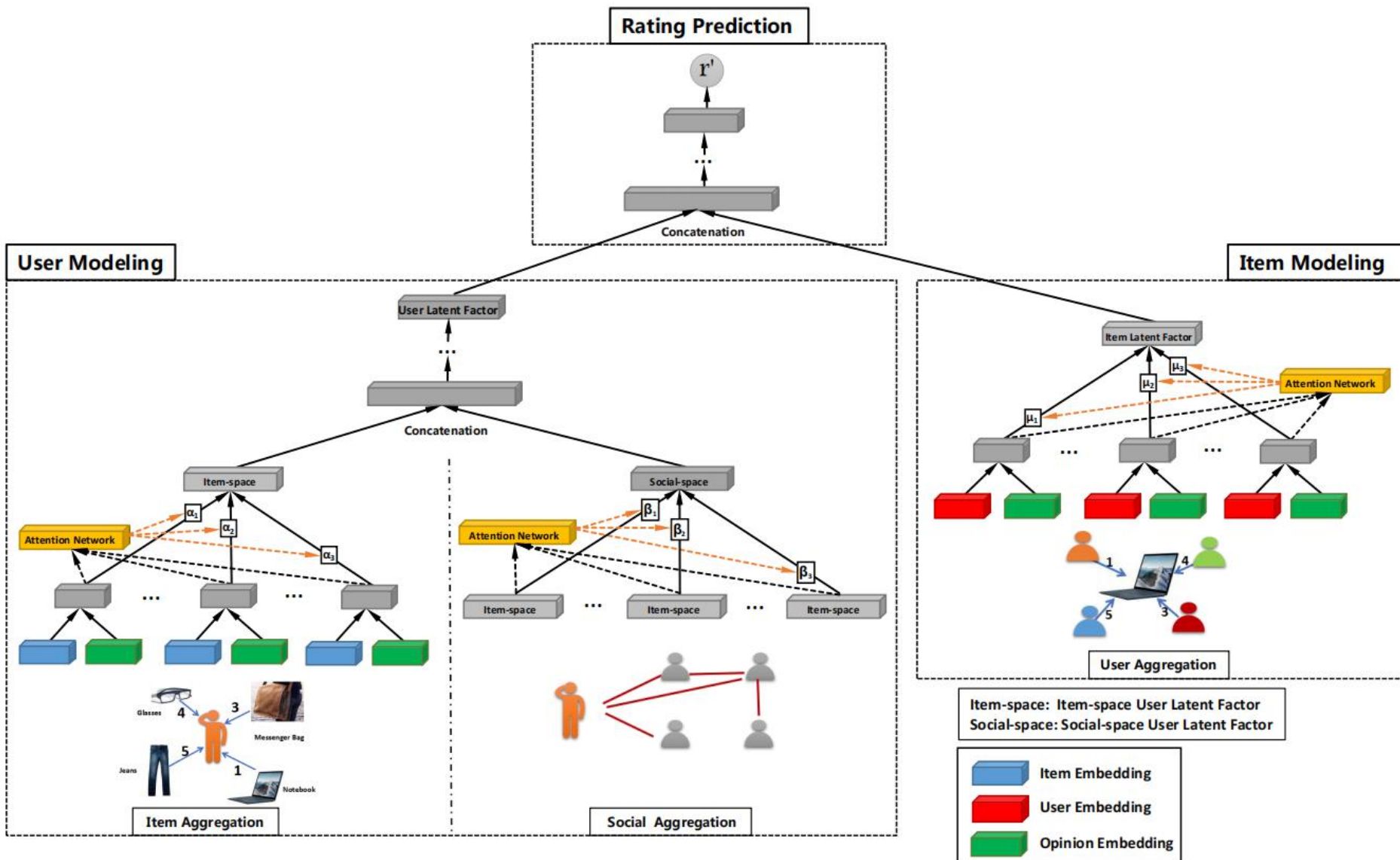
## GraphRec——考虑社交属性（用户和用户之间的交互）



## GraphRec——考虑社交推荐

- 用户建模：等于下面两个建模结果的拼接：
  - ① 由邻居商品聚合到该用户；
  - ② 由邻居用户聚合到该用户。
- 商品建模：由邻居用户聚合到该商品。
- 用户建模和商品建模，用于预测得分。

## GraphRec



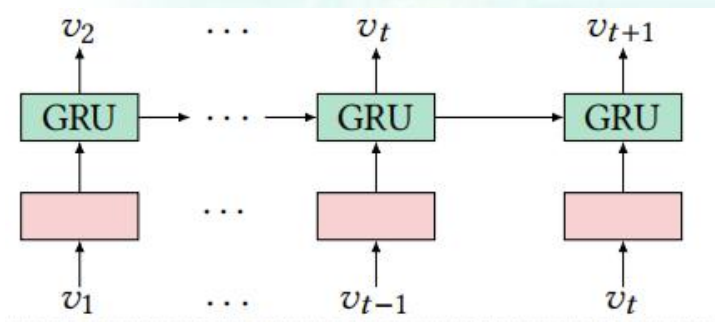
## BERT4Rec——对用户偏好的动态衍变进行建模

- 长期来看，一个用户小时候喜欢买四驱车，长大了可能就喜欢买单反相机；
- 短期来看，一个用户本来想买马桶，但是又看到了其他智能家居，可能就会对智能马桶感兴趣。

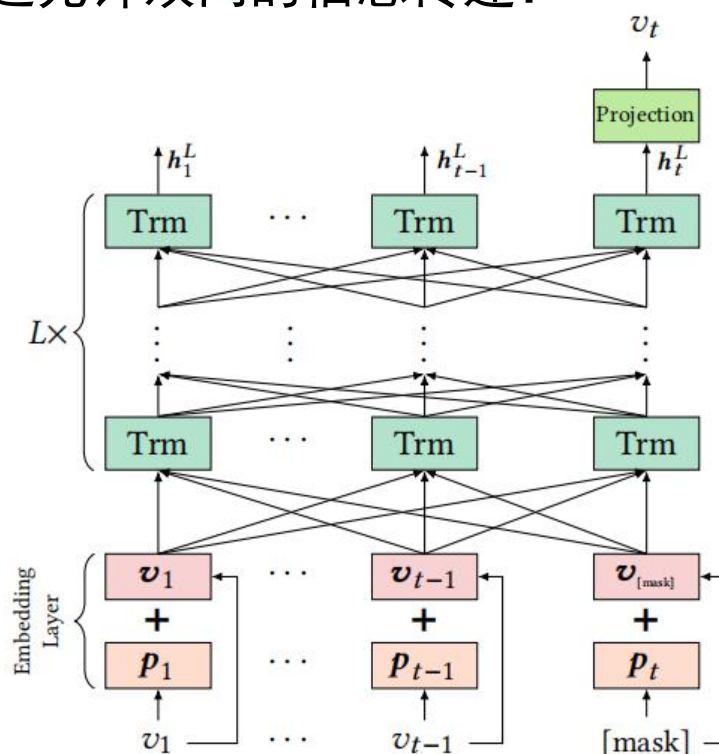


## BERT4Rec——对用户偏好的动态衍变进行建模

相比于普通RNN，BERT4Rec还允许双向的信息传递：



利用普通RNN建模



BERT4Rec