飞桨

图神经网络七日打卡营

小斯妹　百度PGL团队成员

2020.11.27

课程大纲

- Graph Learning
- 图学习

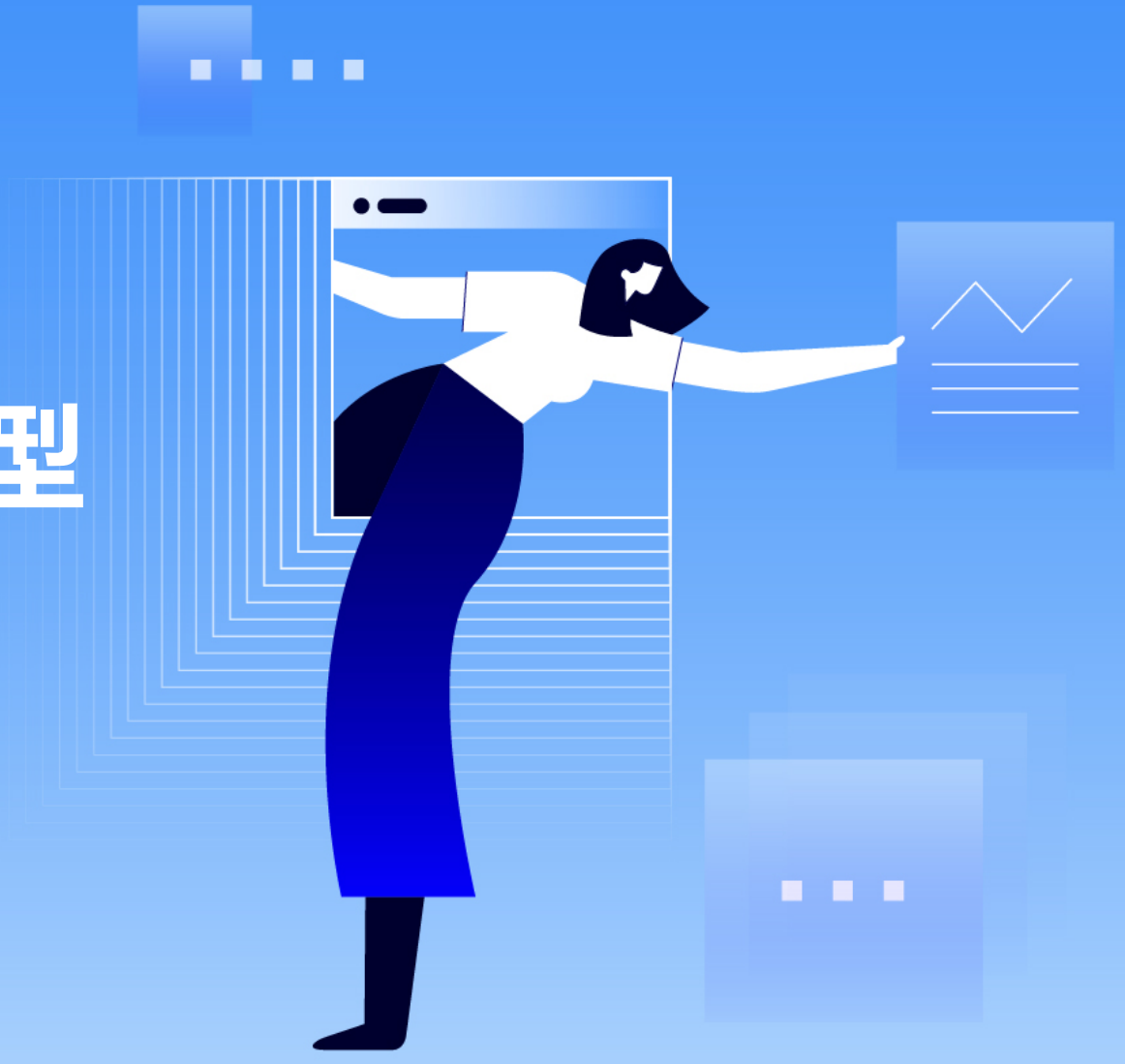
后续：新冠项目实战，带你助力疫情防控

参考材料：

- 斯坦福CS224W课程： http://cs224w.stanford.edu
- 图学习库 PGL：https://github.com/PaddlePaddle/PGL

飞桨

第五课 GNN进阶模型



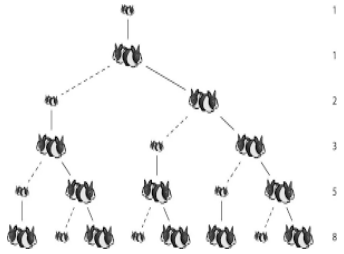
小斯妹 百度PGL团队成员

2020.11.27

简单回顾

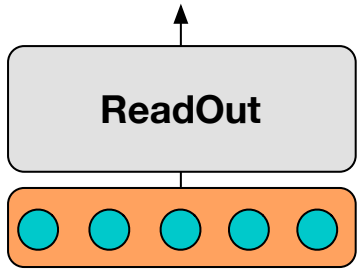
图采样

多阶邻居指数暴涨

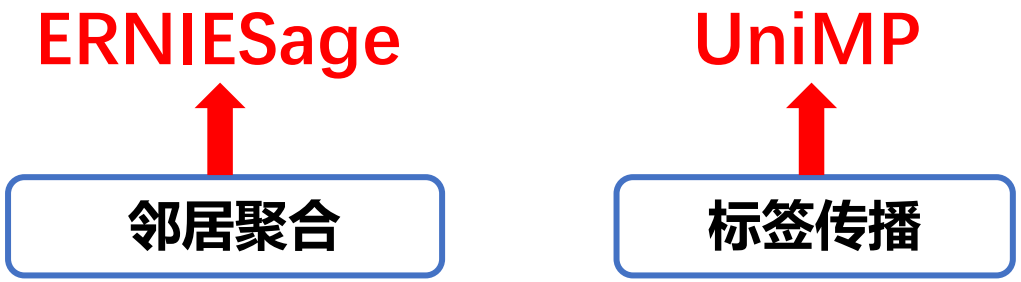


邻居聚合

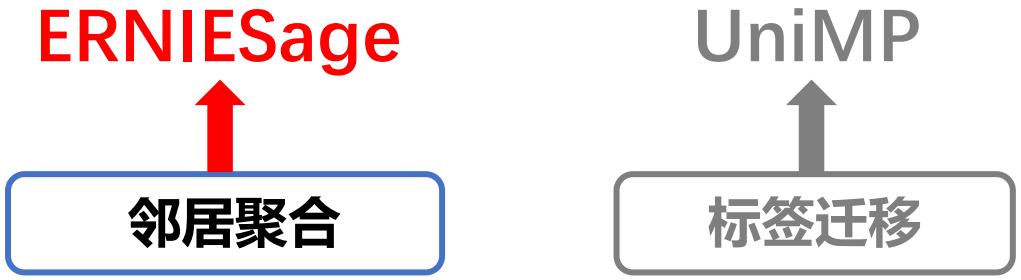
目前的邻居聚合够吗



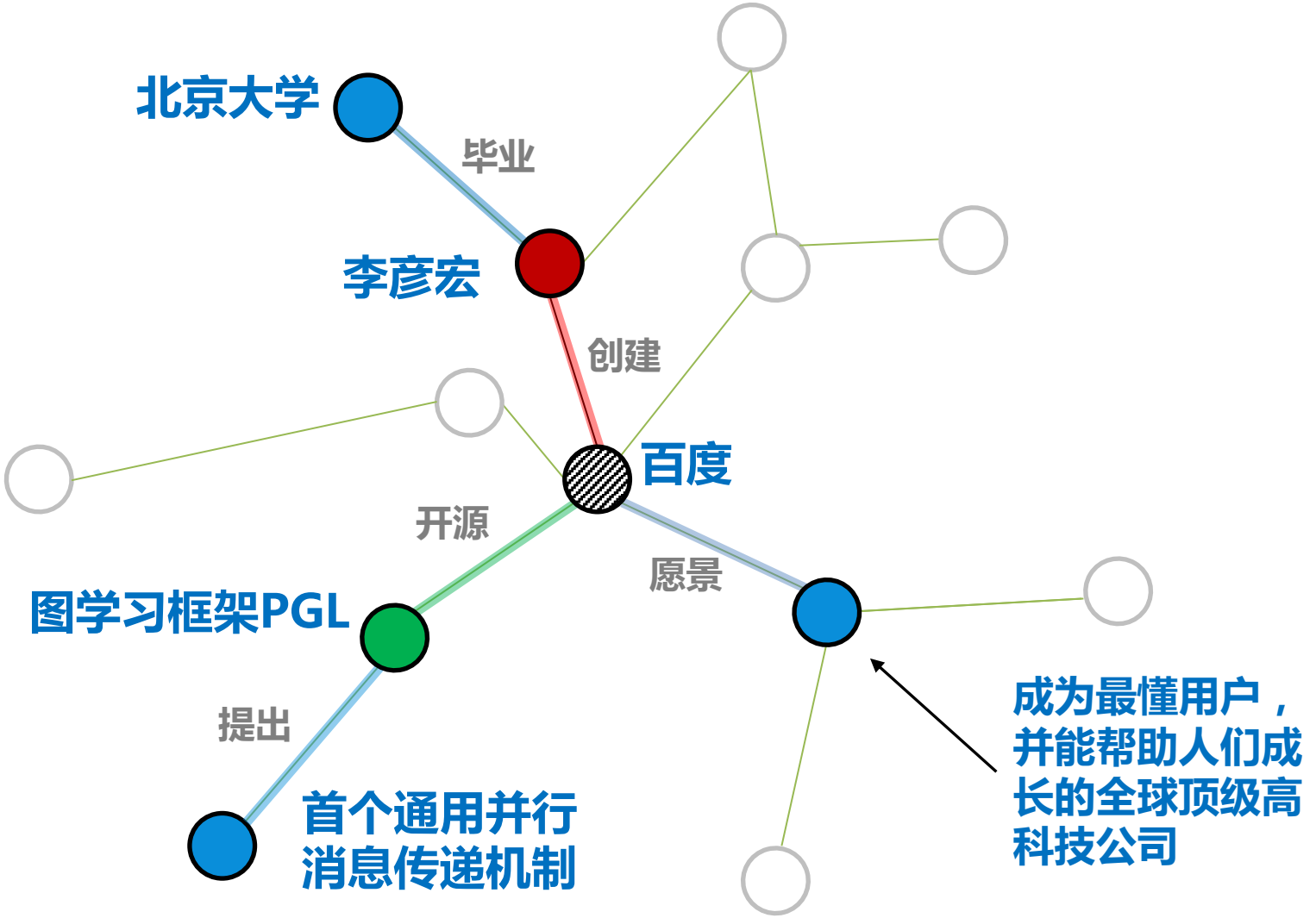
本课内容



本课内容



背景：Text Graph



Text Graph

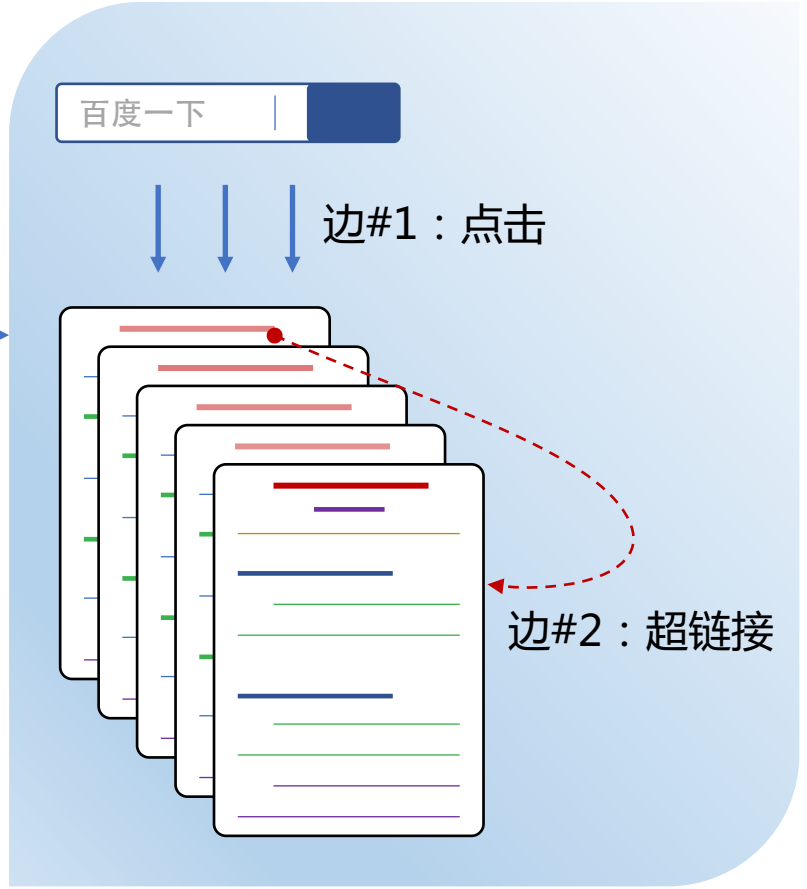
=

节点与边带有文本的特殊图

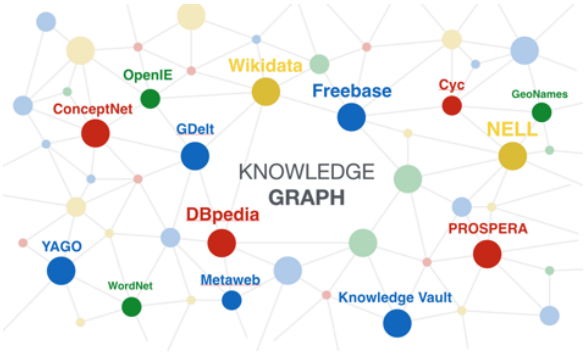
背景：Text Graph广泛存在于工业界应用中

信息流推荐

搜索场景



贴吧推荐



知识图谱

如何对Text Graph进行建模？

只关注结构信息

GraphSage

退化成文本理解

ERNIE

百度推出的语义理解技术

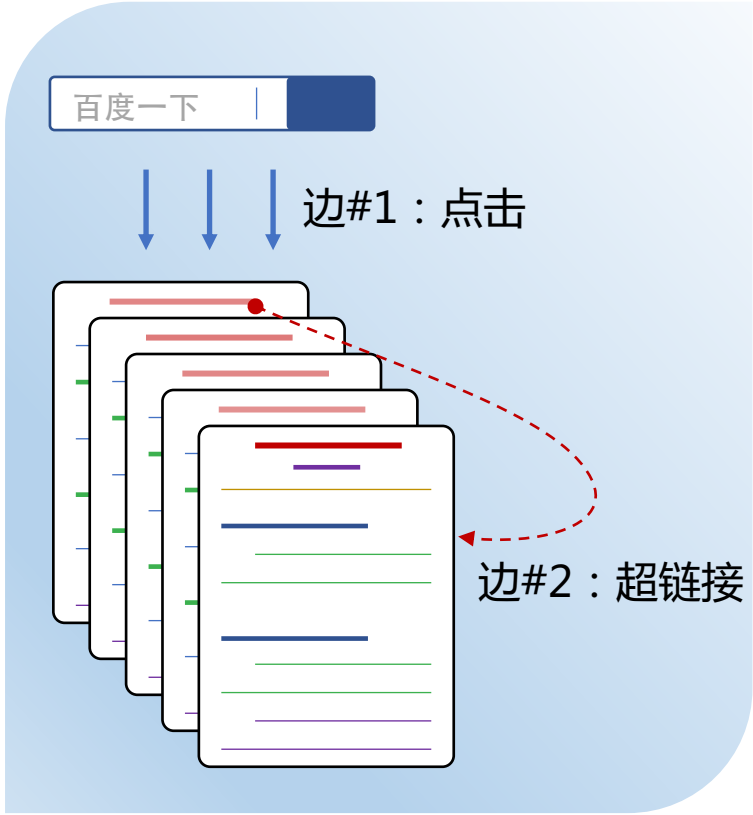
对Text Graph进行建模已有方法的局限

Introducing ERNIESage

https://github.com/PaddlePaddle/PGL/tree/main/examples/erniesage

图语义理解技术ERNIESage

结构建模



语义理解

Introducing ERNIESage (cont.)

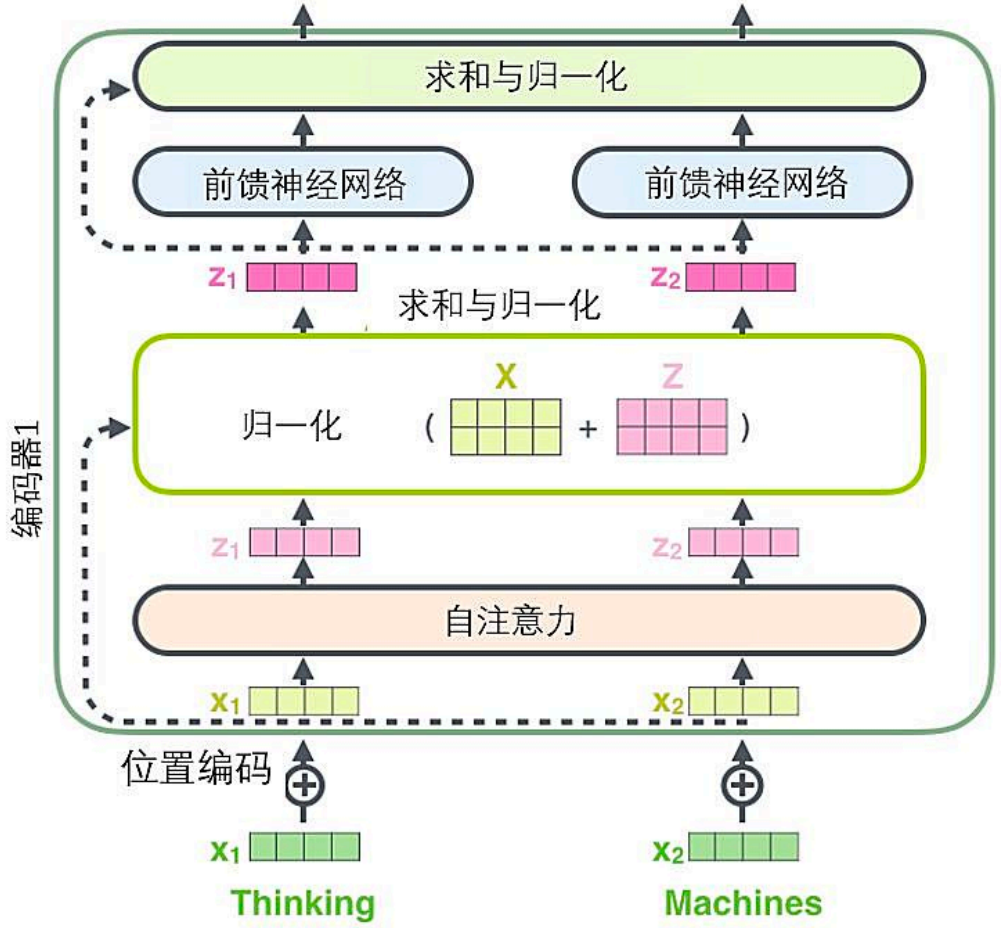
ERNIESage: ERNIE SAmple aggreGatE

图语义理解技术ERNIESage = ERNIE + GraphSage

在第四节

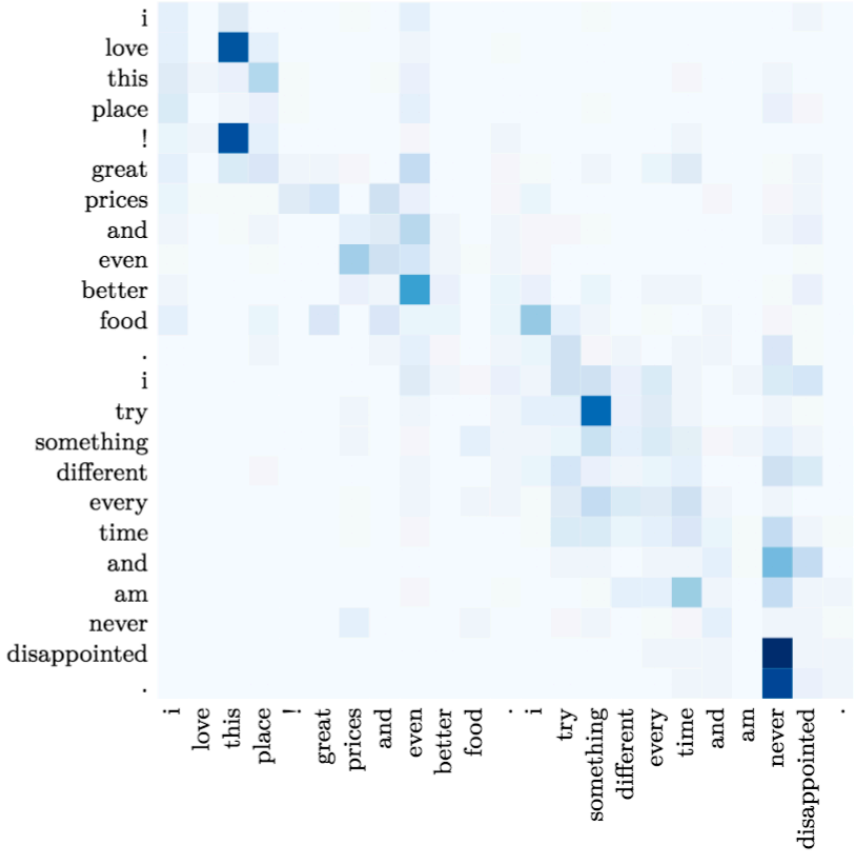
已经介绍过了☺

ERNIE背景知识介绍：Transformer基础结构



Transformer编码器

《Attention Is All You Need》

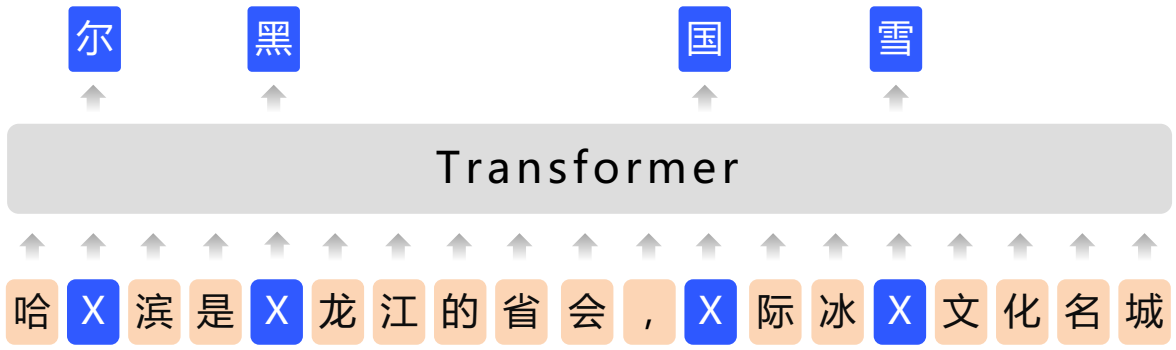


Transformer自注意力机制发现搭配关系

ERNIE 1.0核心思想：知识增强

Masked Language Model(MLM)

Learnt by **BERT**



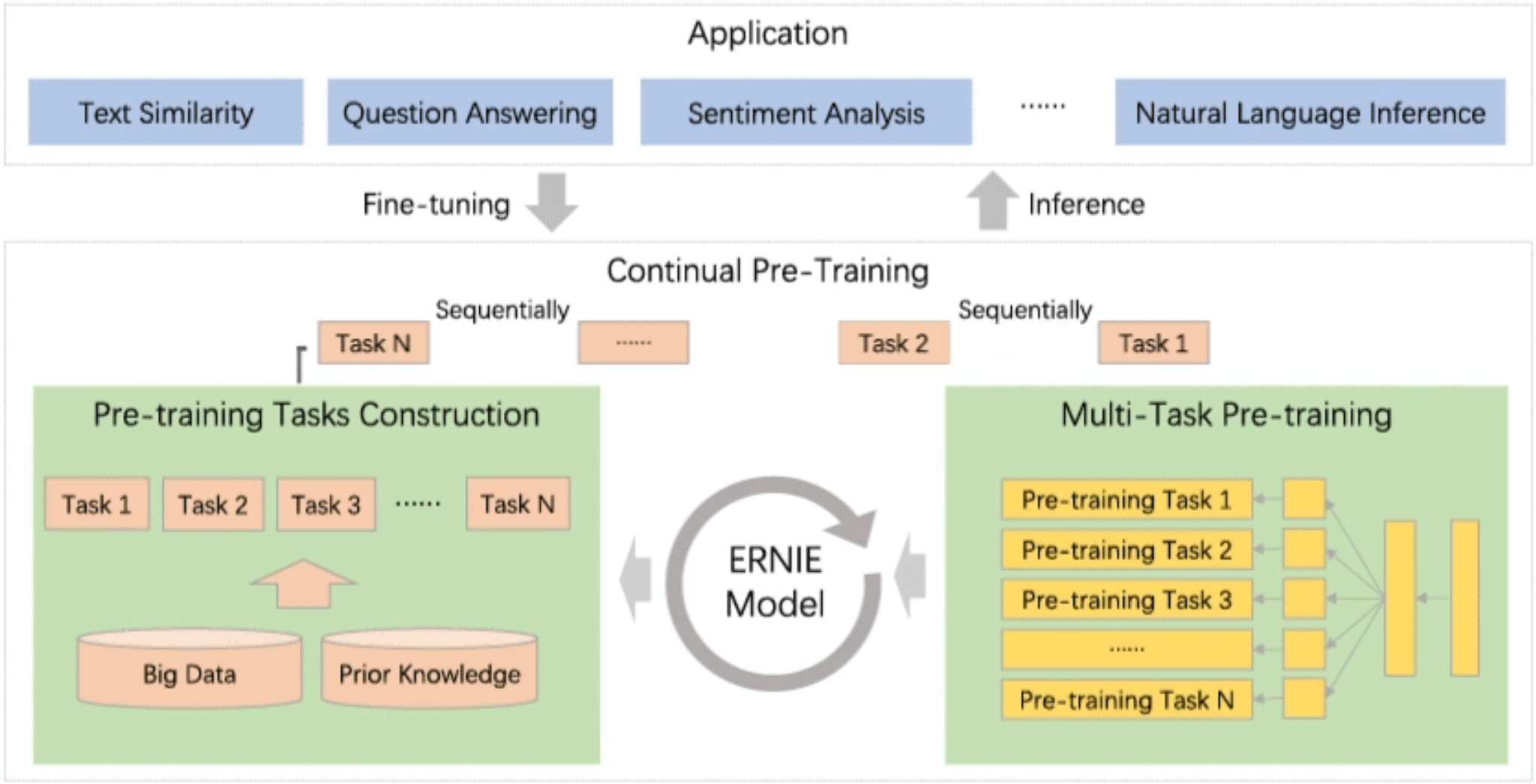
哈尔滨是黑龙江的省会，国际冰雪文化名城

Learnt by **ERNIE**

哈尔滨是黑龙江的省会，国际冰雪文化名城

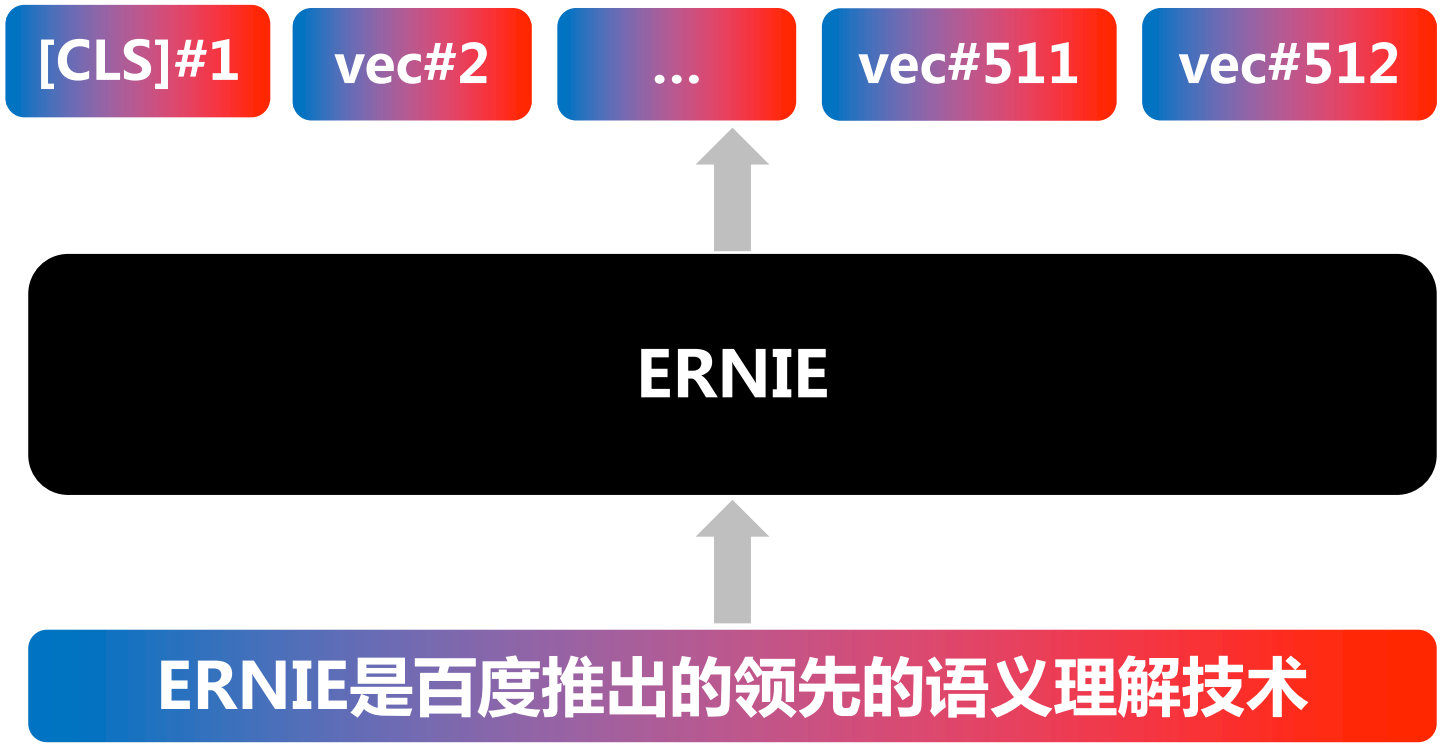
BERT: https://arxiv.org/abs/1810.04805, ERNIE1.0: https://arxiv.org/abs/1904.09223

ERNIE 2.0核心思想：持续学习



ERNIE 2.0：持续学习知识效果

ERNIE小结



ERNIE是百度推出的语义理解技术

1. 世界领先的性能
2. 广泛的应用前景

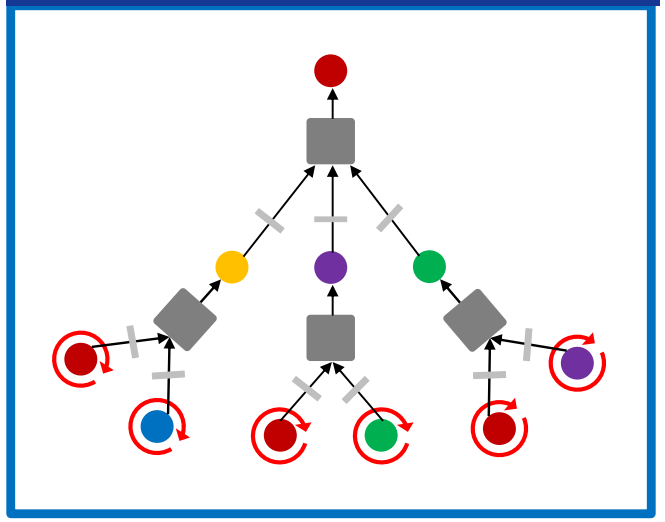
- **我们只需黑盒使用即可，像这样：**

```
from models.ernie_model.ernie import ErnieModel
ernie = ErnieModel(src_ids=text_ids...)
```

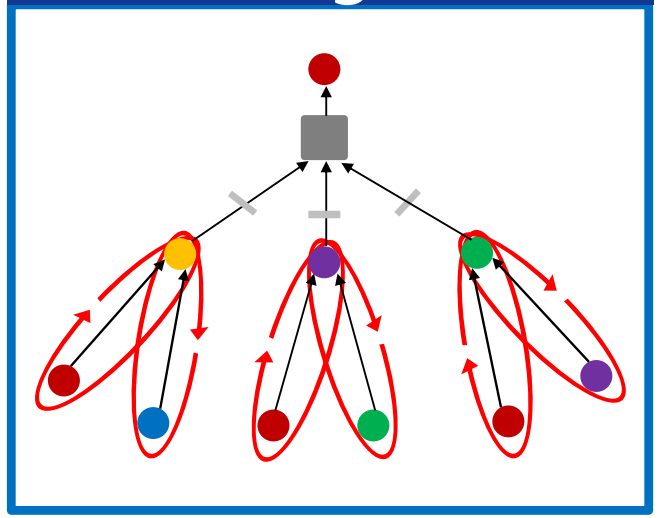
让我们回到ERNIESage

将ERNIE作用于Graph的各大元素

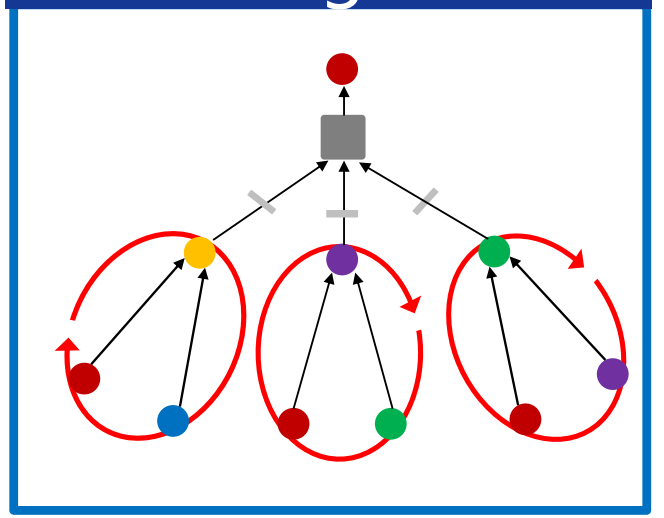
ERNIESage Node



ERNIESage Edge

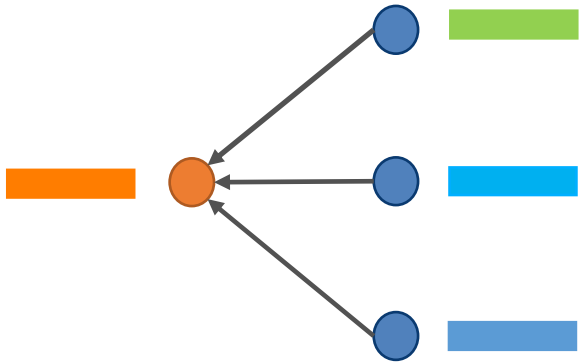


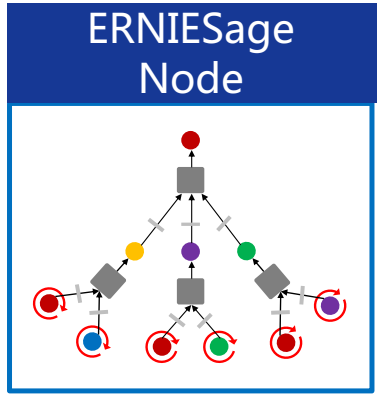
ERNIESage 1-Neighbor



ERNIESage Node

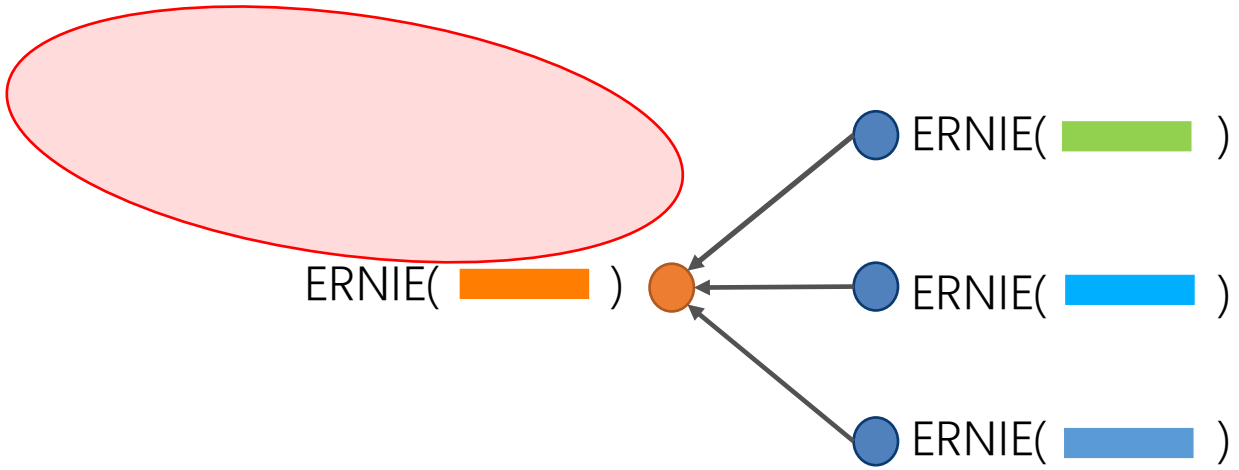
ERNIE 作用于Text Graph的Node（节点）上

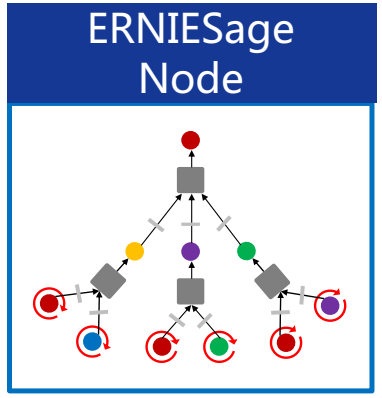




ERNIESage Node

ERNIE 作用于Text Graph的Node（节点）上

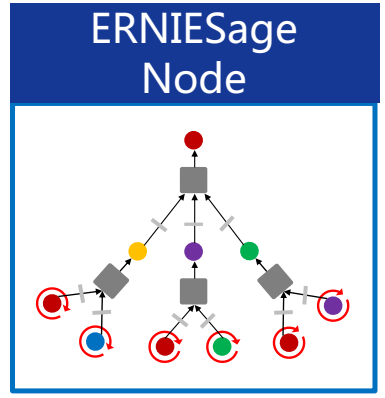


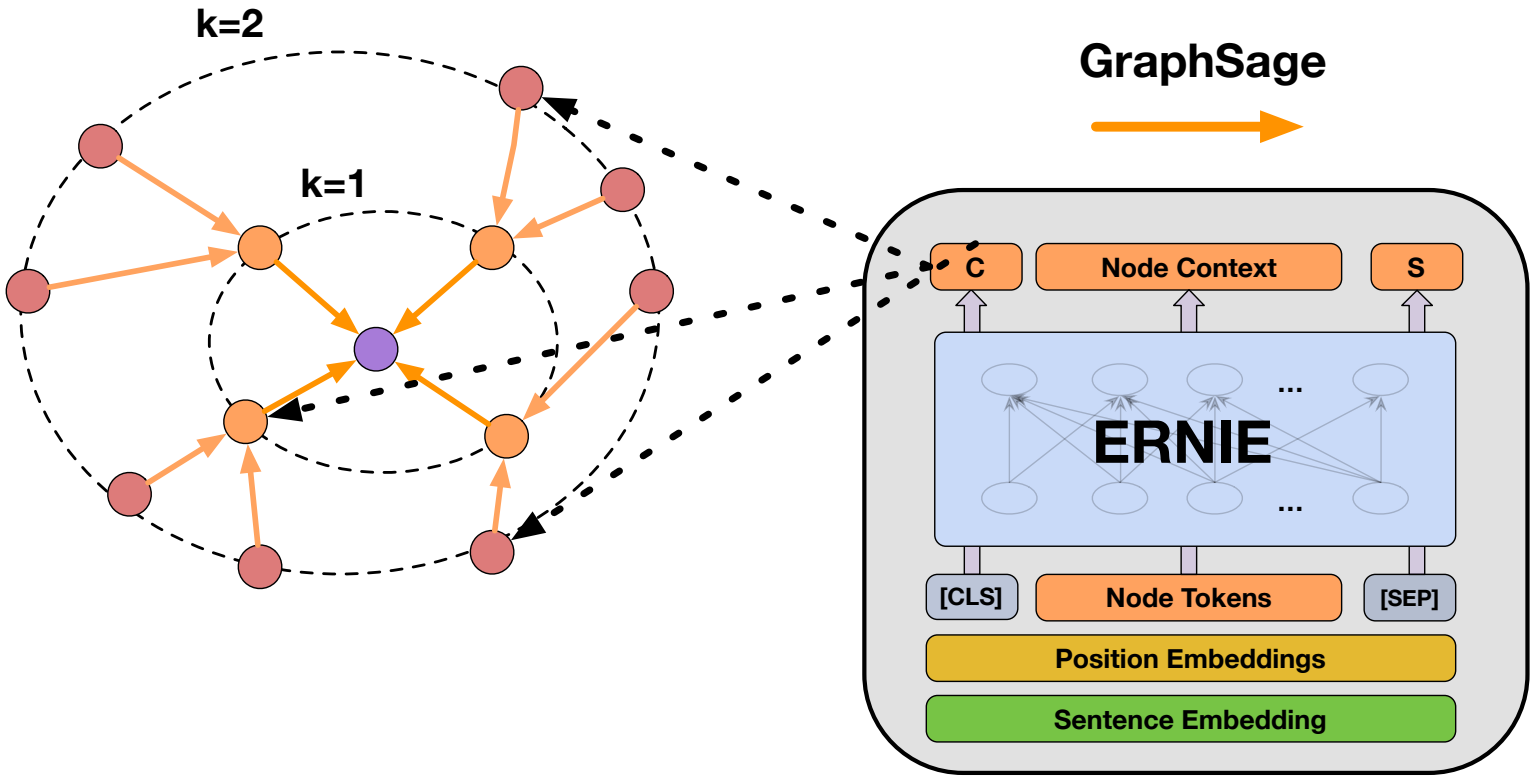


1. 利用 ERNIE 获取节点表示
2. 聚合邻居特征：ERNIE( ) + ERNIE( ) + ERNIE( ) Send, Recv
3. 将当前节点和聚合后的邻居特征 concat，更新节点特征 Update

ERNIESage Node：模型结构

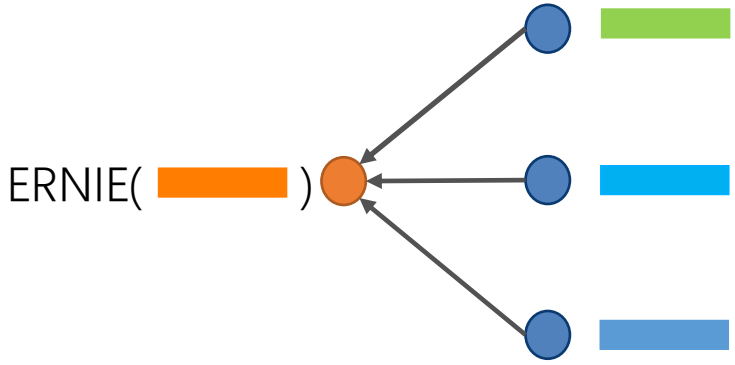
ERNIE 作用于Text Graph的Node（节点）上

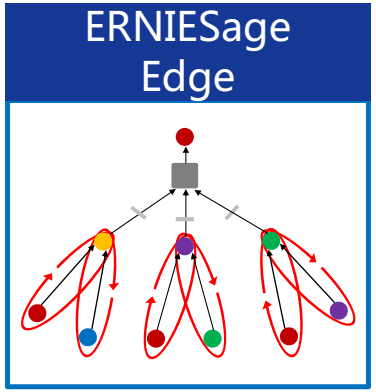




ERNIESage Edge

ERNIE聚合Text Graph的Edge（边）上信息

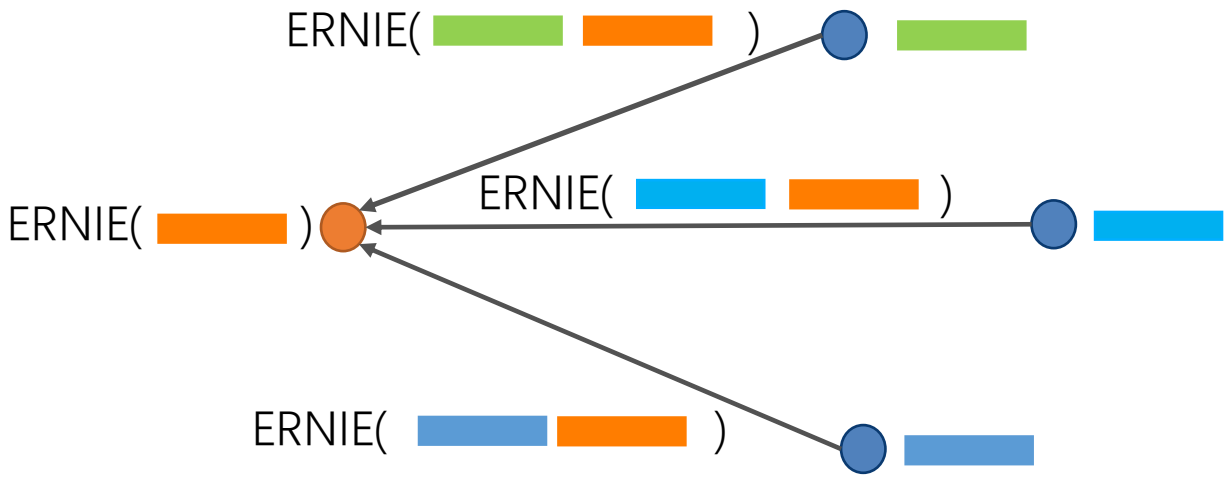


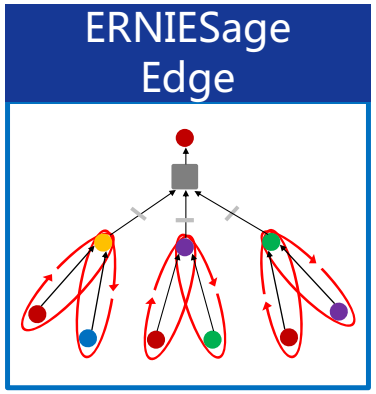


1. 利用 ERNIE 获取中心节点的文本特征表示

ERNIESage Edge

ERNIE聚合Text Graph的Edge（边）上信息





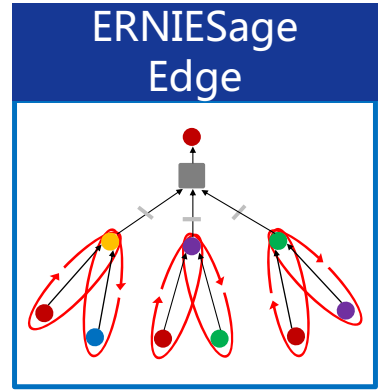
1. 利用 ERNIE 获取中心节点的文本特征表示
2. **交互**：中心节点和邻居节点一同作为 ERNIE 输入，计算交互特征

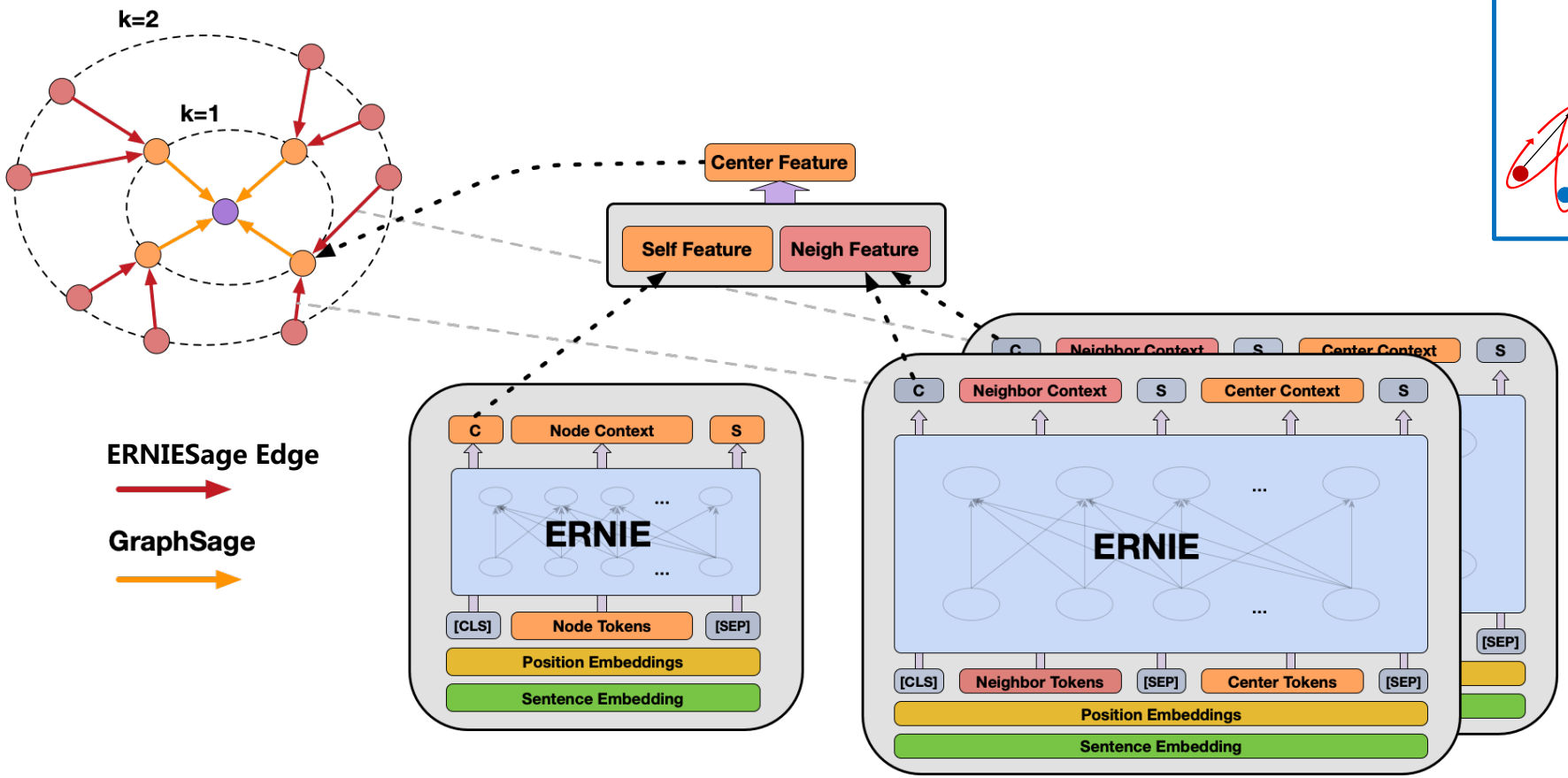
Send, Recv

3. 聚合邻居特征：ERNIE( ) + ERNIE( ) + ERNIE( )
4. 将当前节点和聚合后的邻居特征 concat，更新节点特征 Update

ERNIESage Edge：模型结构

ERNIE聚合Text Graph的Edge（边）上信息

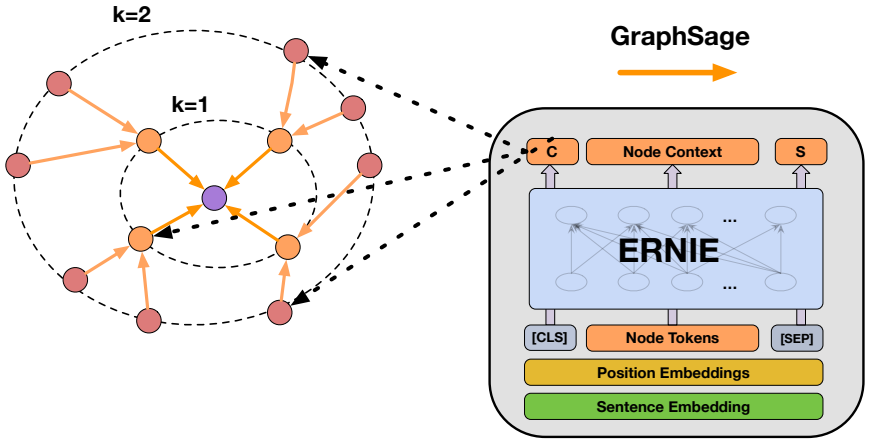




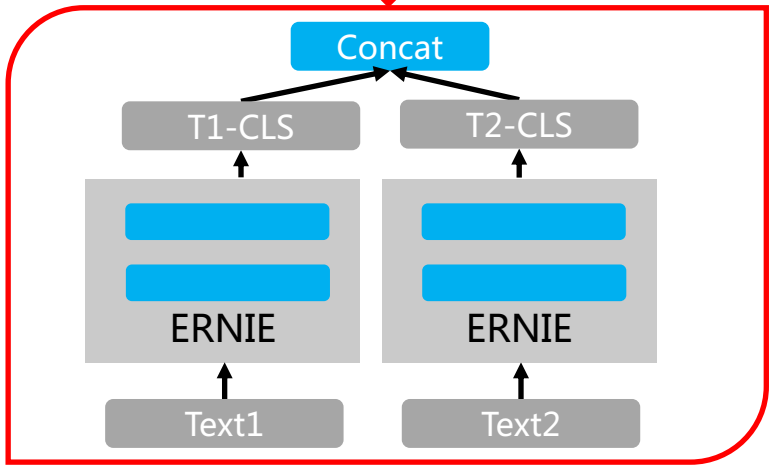
输入：边的 (src, dst)

ERNIESage Node与ERNIESage Edge对比

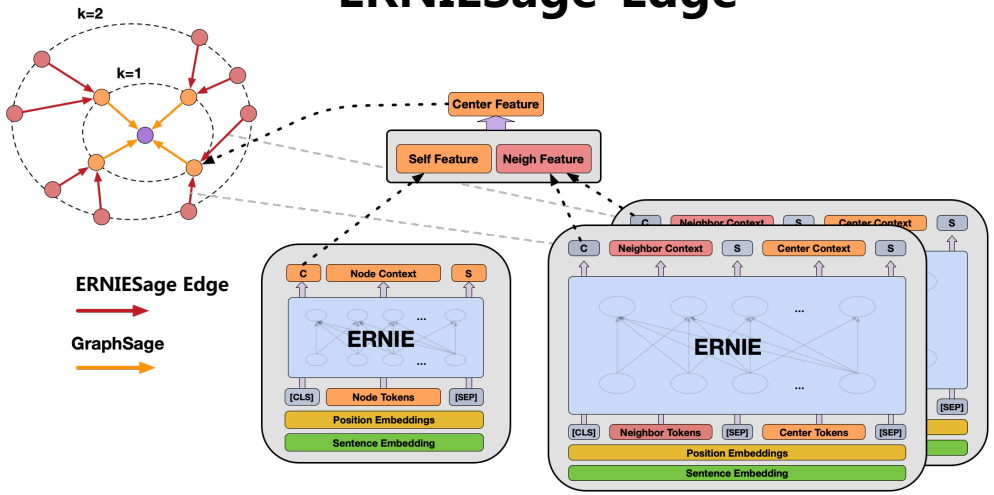
ERNIESage-Node



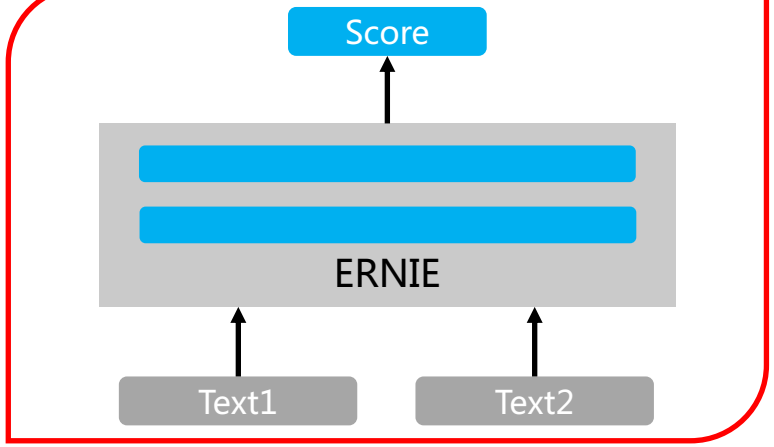
双塔模型



ERNIESage-Edge



单塔模型



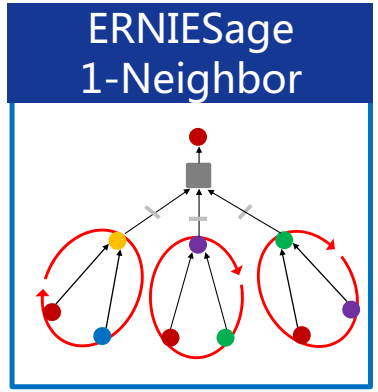
<

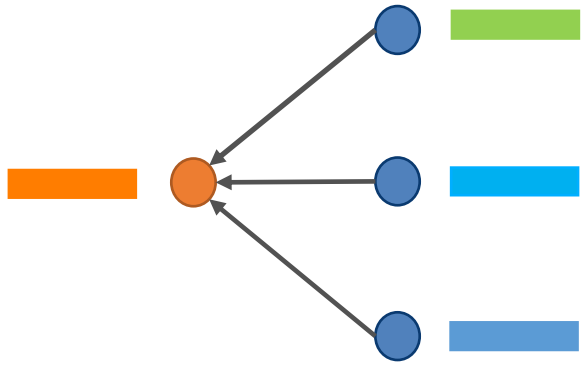
大部分场景验证

单塔效果好于双塔

ERNIESage 1-Neighbor

ERNIE聚合节点的1 Neighbor（**一阶邻居**）信息

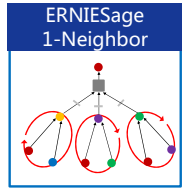




ERNIE()

将中心节点的文本与所有一阶邻居节点文本进行**单塔拼接**，再利用 ERNIE 做消息聚合。

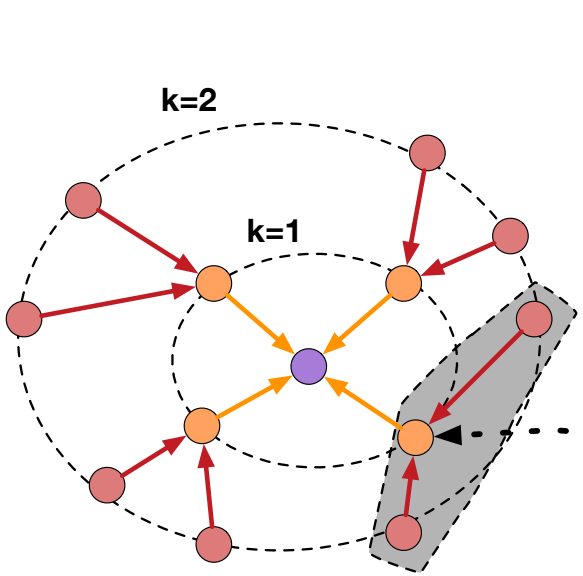
ERNIESage 1-Neighbor：模型结构



ERNIE聚合节点的1 Neighbor（**一阶邻居**）信息

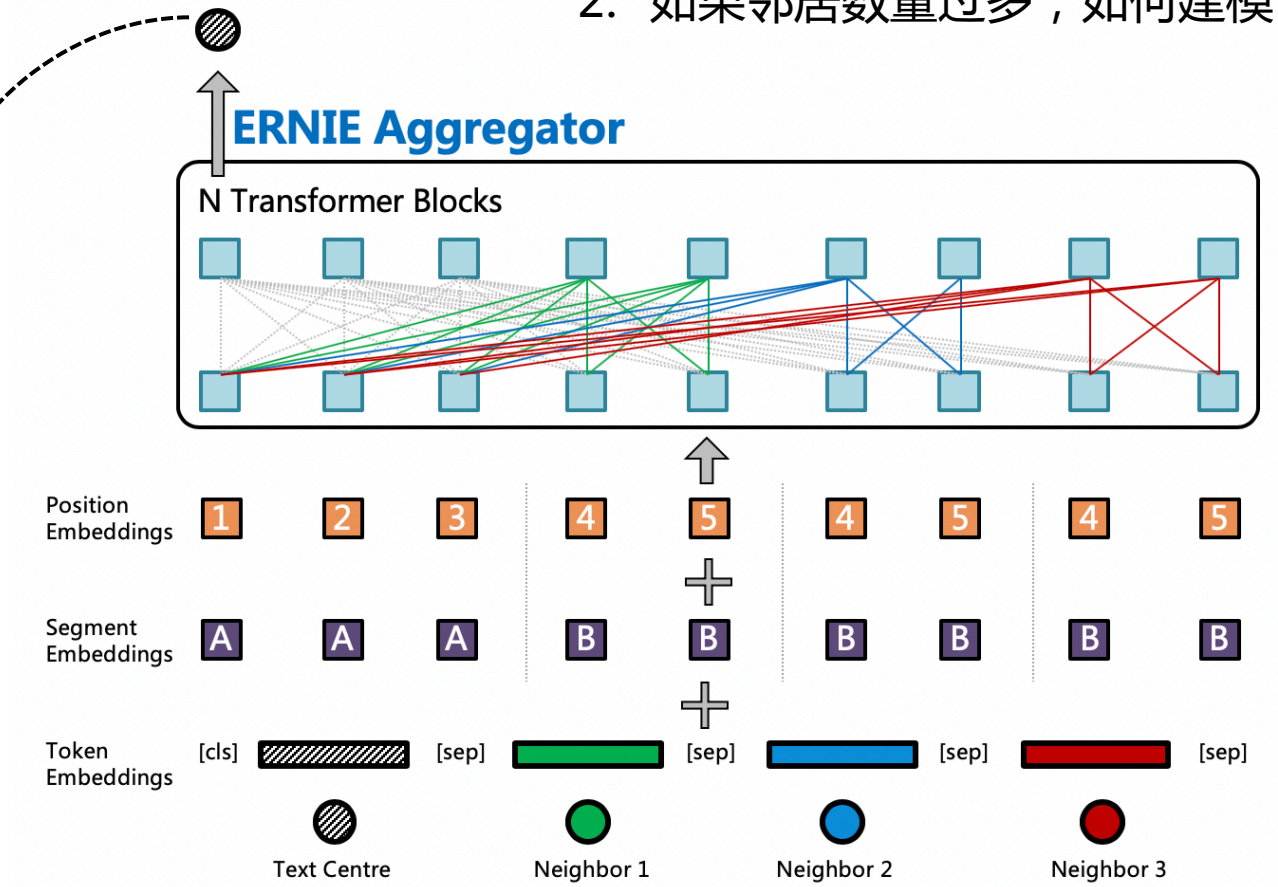
问题：

1. 如何确保在输入时邻居不带有顺序？
2. 如果邻居数量过多，如何建模？

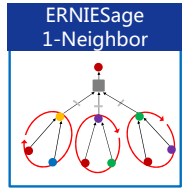


ERNIESage 1-Neighbor

GraphSage



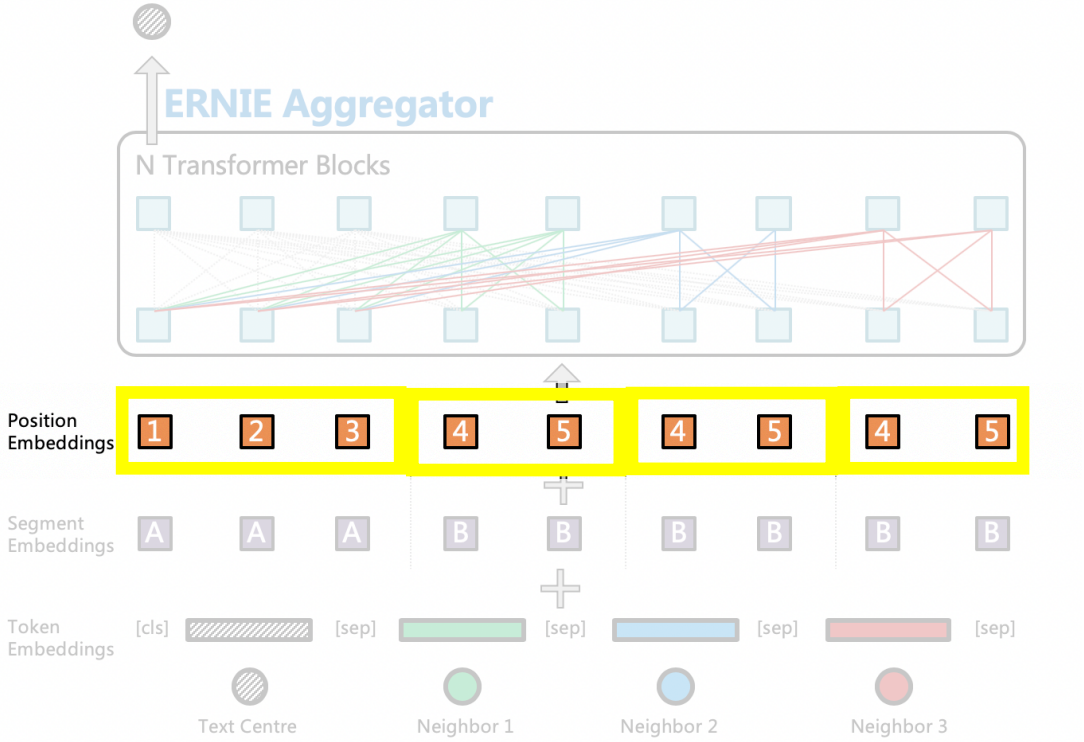
ERNIESage 1-Neighbor：模型结构



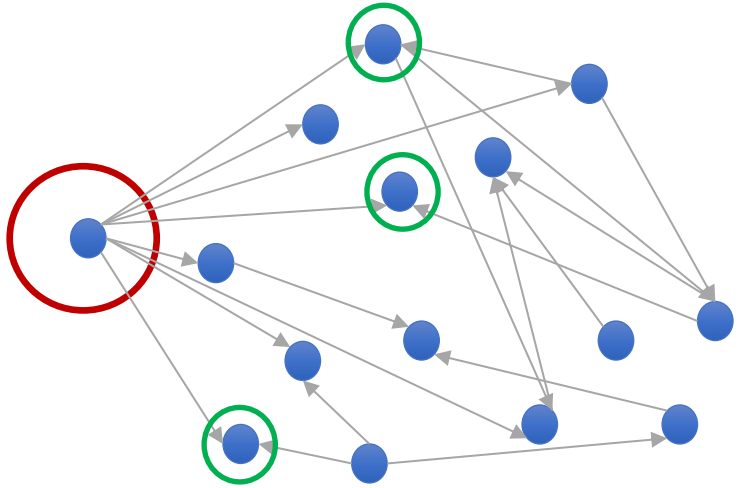
问题：

1. 如何确保在输入时邻居不带有顺序？
2. 如果邻居数量过多，如何建模？

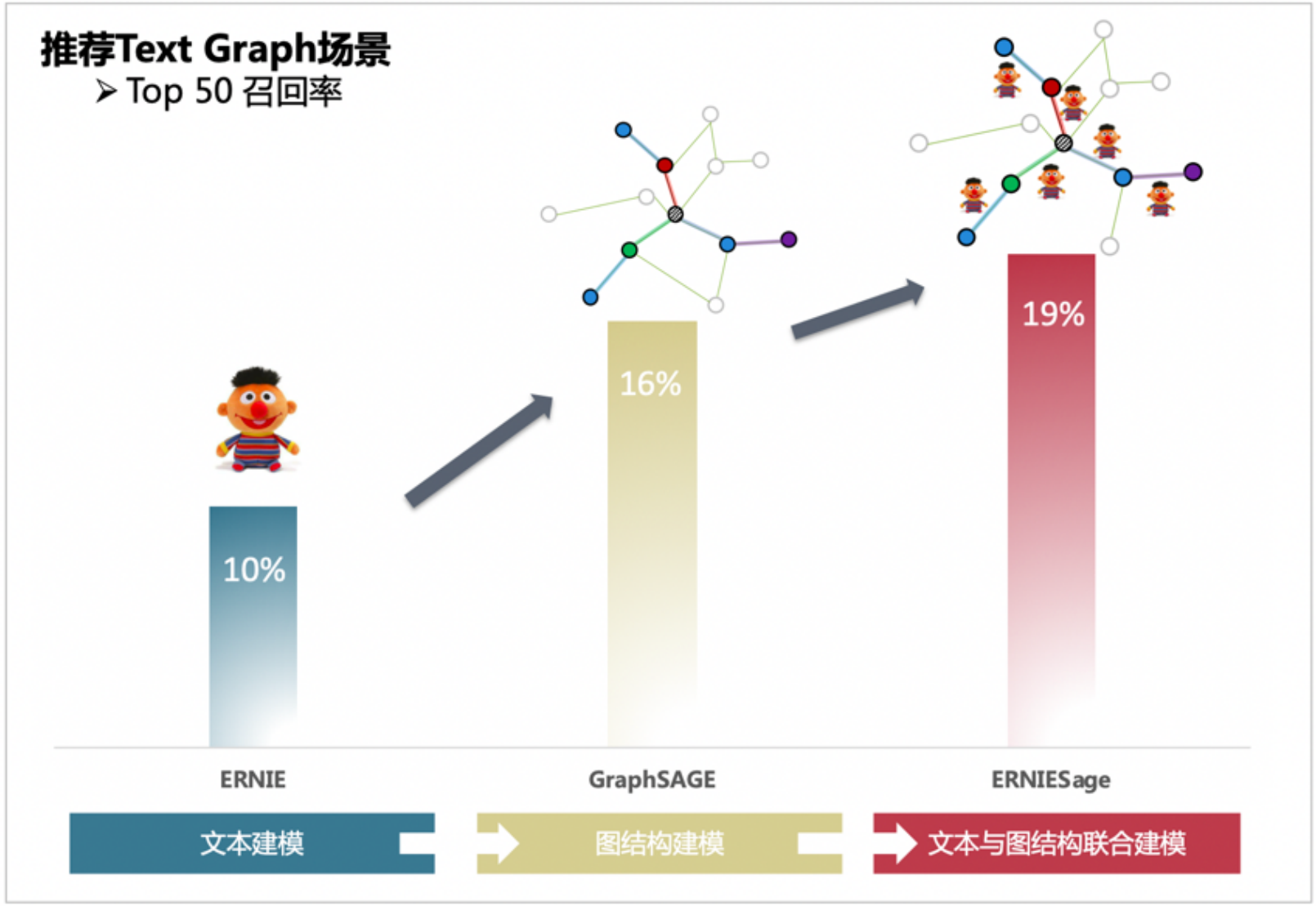
针对问题1：独立的位置编码



针对问题2：对邻居进行采样



ERNIESage：百度推荐场景落地



ERNIESage：

COLING TextGraphs-14 竞赛排名第一

CodaLab Search Competitions My Co

Competition

Multi-Hop Inference Explanation Regeneration (TextGraphs-14)

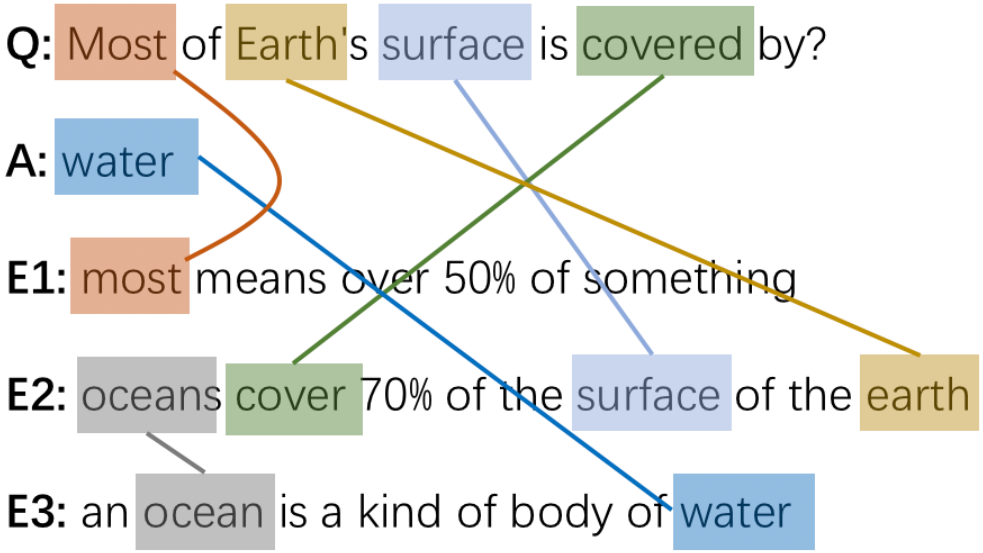
Organized by dustalov - Current server time: Oct. 23, 2020, 3:05 a.m. UTC

Previous	▶ Current	End
Evaluation	Post-Competition	Competition Ends
April 6, 2020, midnight UTC	Sept. 22, 2020, midnight UTC	Never

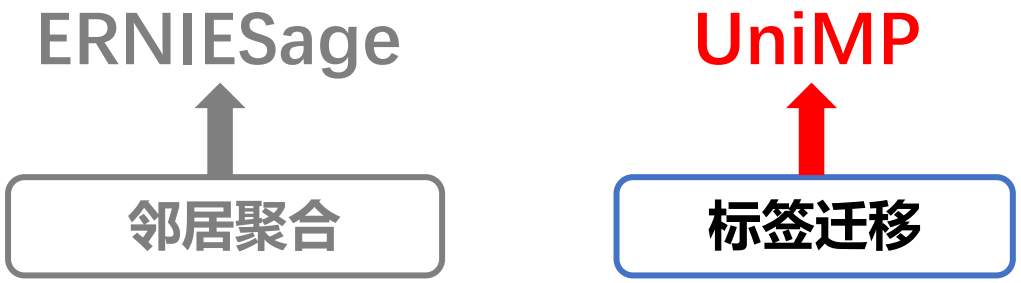
#	User	Entries	Date of Last Entry	Team Name	MAP ▲
1	webbley	26	09/21/20	Baidu PGL	0.6033 (1)
2	alvysinger	31	06/25/20	LIIR - KU Leuven	0.5843 (2)
3	aisys	9	07/10/20		0.5233 (3)
4	AGP	18	09/21/20	ChiSquareX	0.4902 (4)
5	vmm	10	09/21/20		0.4902 (4)
6	redken	3	09/21/20	Red Dragon	0.4793 (5)
7	Aditya-Pawate	2	09/19/20	Team IITian	0.4519 (6)
8	mler	2	09/07/20		0.3367 (7)

任务目标：在**解释图**中进行多步文本推理，召回出符合推理过程的若干节点

Multi-hop Inference by Explanation Graph



本课内容



OGB 打榜神器：UniMP

UniMP: 融合标签传递和图神经网络的统一模型

Leaderboard for ogbn-products

The classification accuracy on the test and validation sets. The higher, the better.

Package: >=1.1.1

Rank	Method	Test Accuracy	Validation Accuracy	Contact	References	#Params
1	**UniMP**	0.8256 ± 0.0031	0.9308 ± 0.0017	Yunsheng Shi (PGL team)	Paper, Code	1,475,605

Leaderboard for ogbn-arxiv

The classification accuracy on the test and validation sets. The higher, the better.

Package: >=1.1.1

Rank	Method	Test Accuracy	Validation Accuracy	Contact	References	#Params
1	**UniMP_large**	0.7379 ± 0.0014	0.7475 ± 0.0008	Yunsheng Shi (PGL team)	Paper, Code	1,162,515

Leaderboard for ogbn-proteins

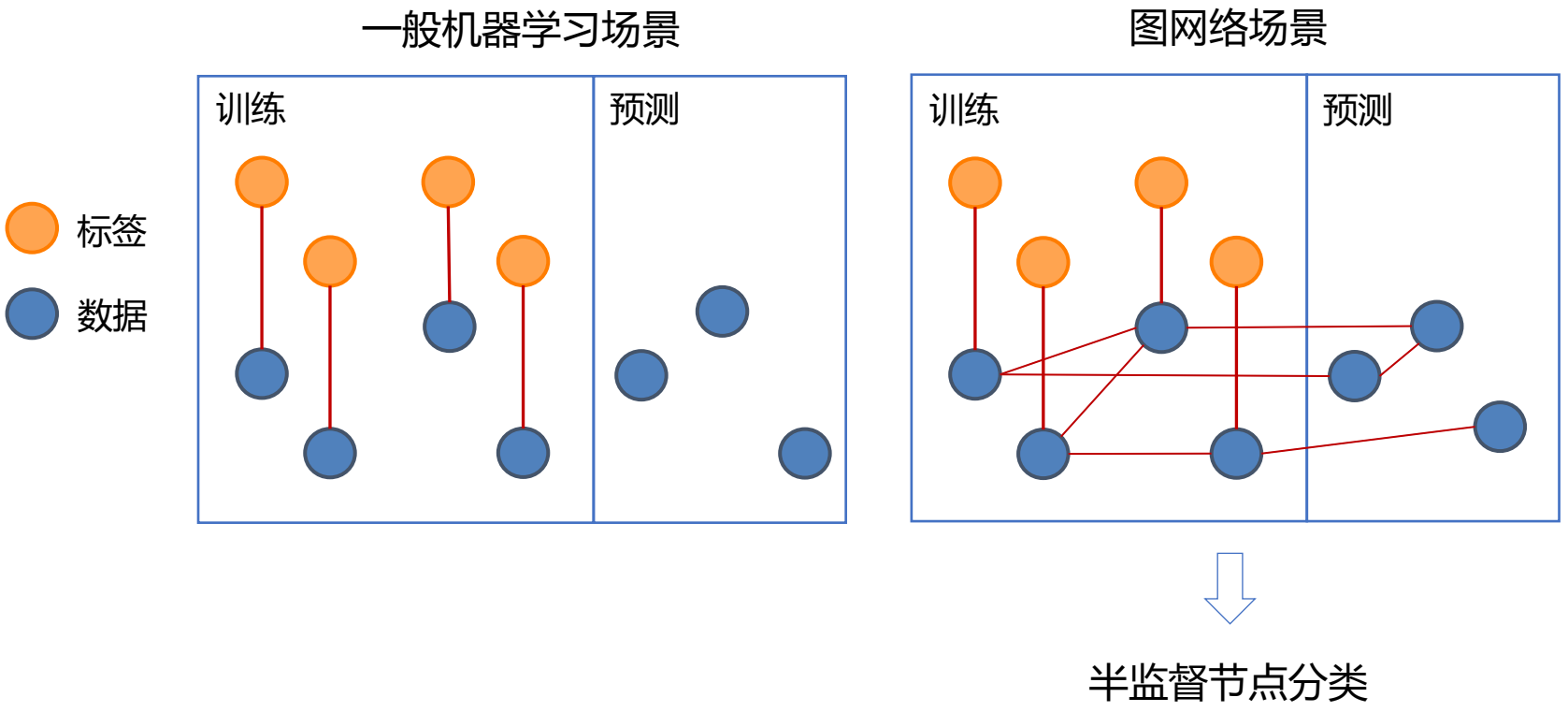
The ROC-AUC score on the test and validation sets. The higher, the better.

Package: >=1.1.1

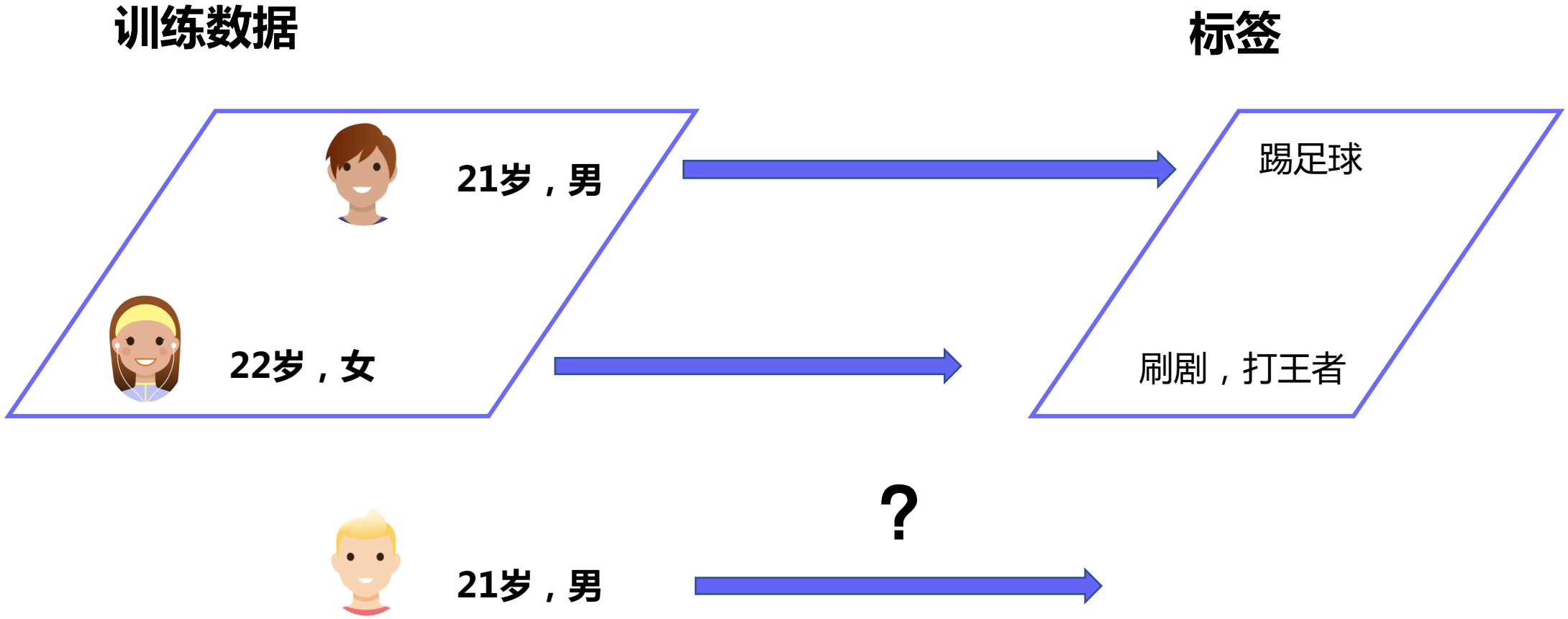
Rank	Method	Test ROC-AUC	Validation ROC-AUC	Contact	References	#Params
1	**UniMP**	0.8642 ± 0.0008	0.9175 ± 0.0006	Yunsheng Shi (PGL team)	Paper, Code	1,909,104

MASKED LABEL PREDICTION: UNIFIED MESSAGE PASSING MODEL FOR SEMI-SUPERVISED CLASSIFICATION

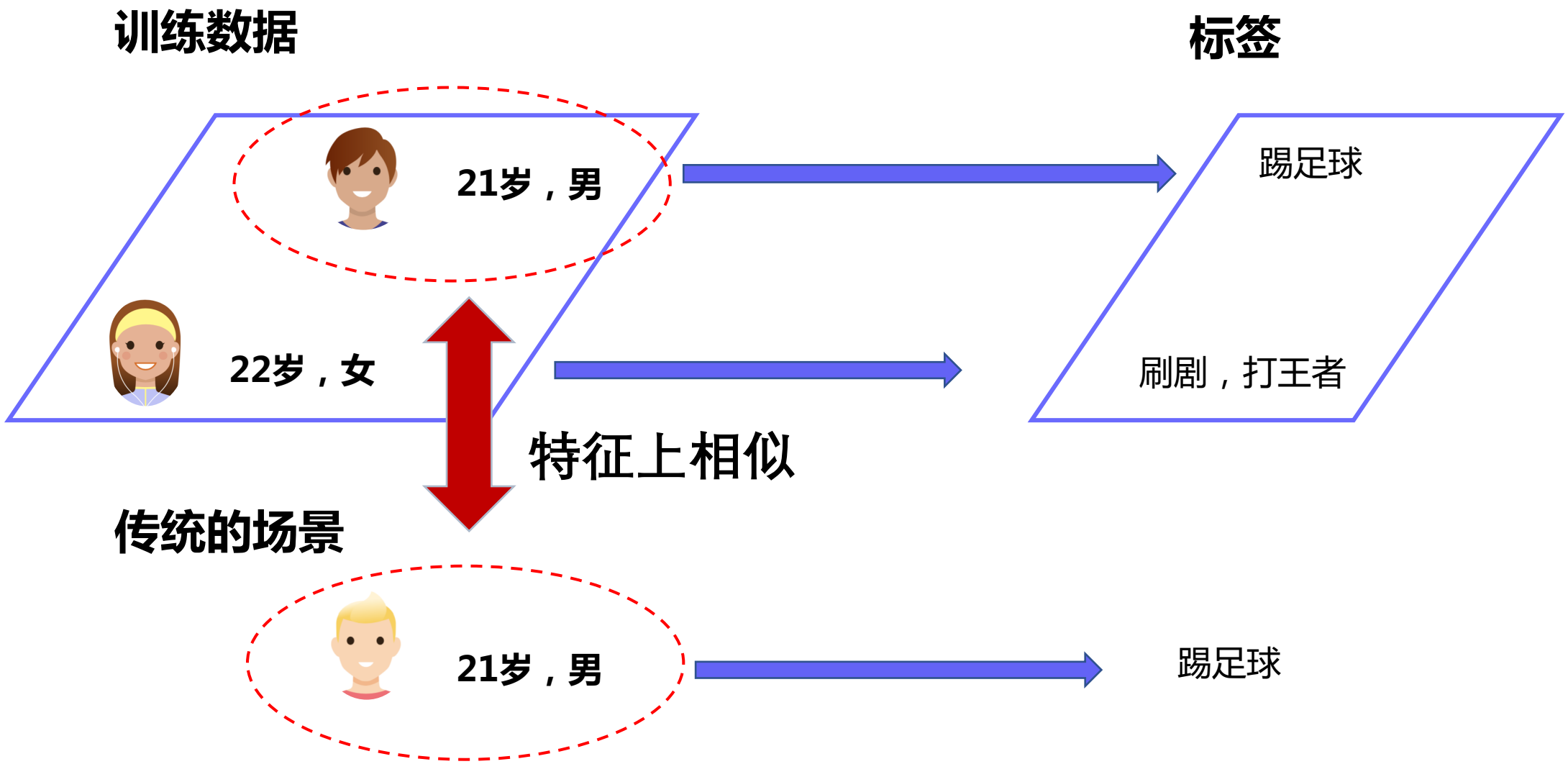
提出背景



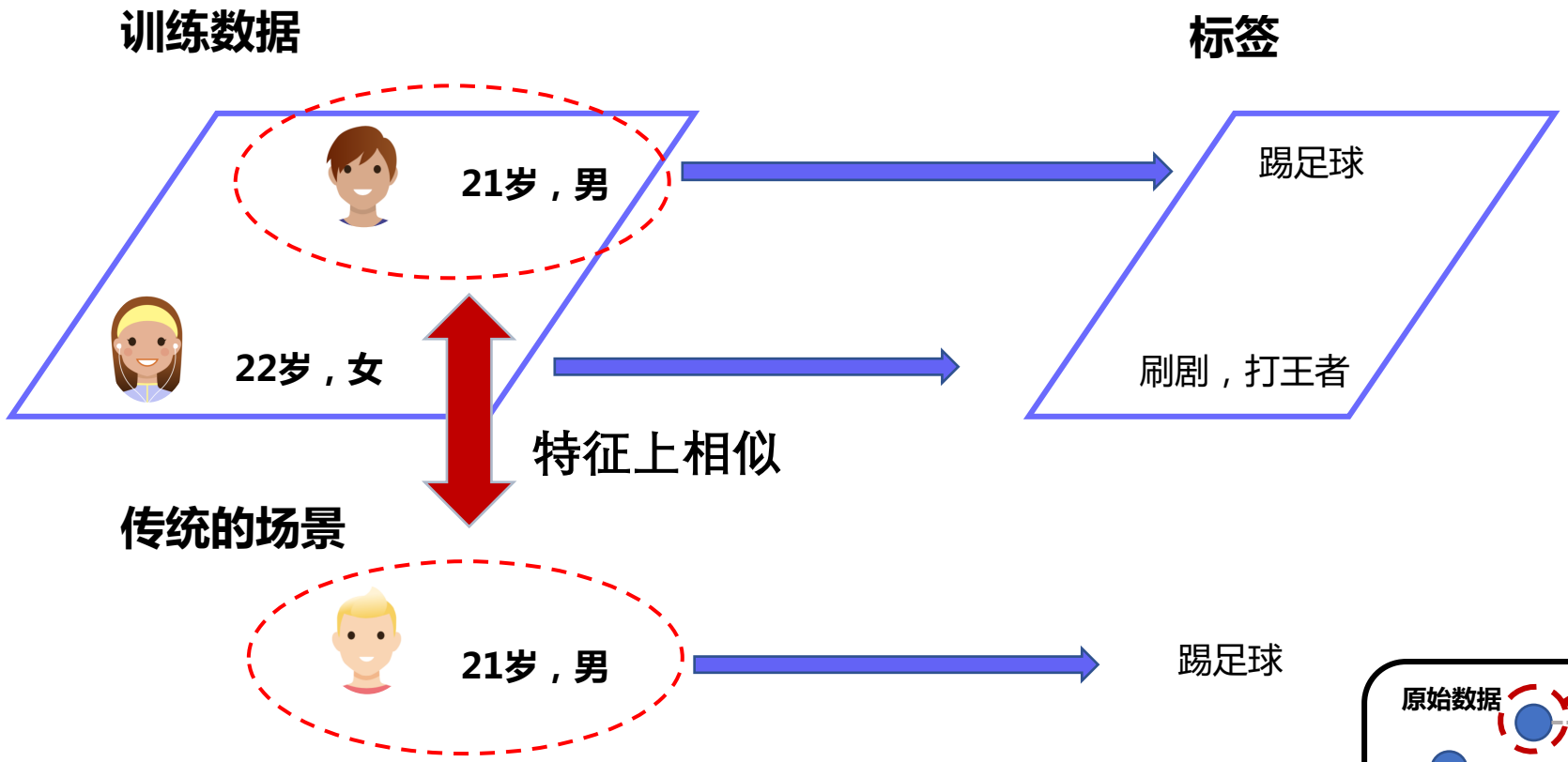
提出背景



提出背景

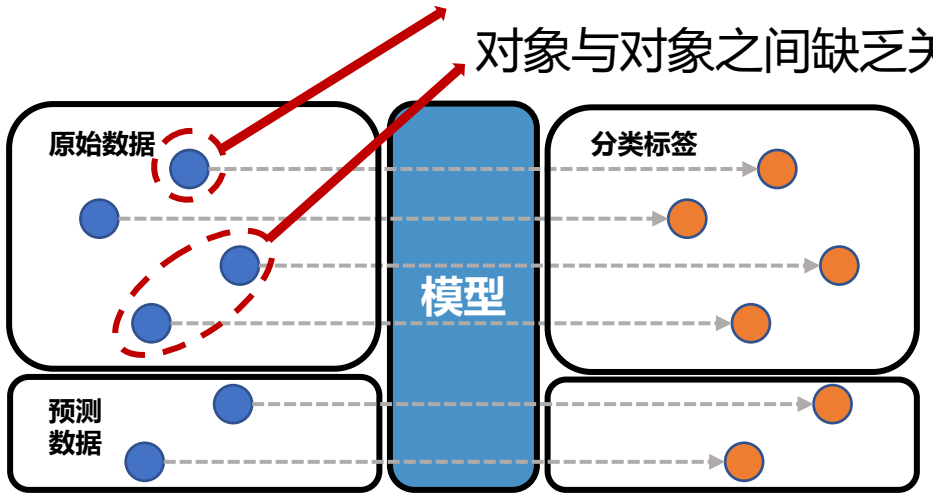


提出背景



研究的对象独立：图像、文本

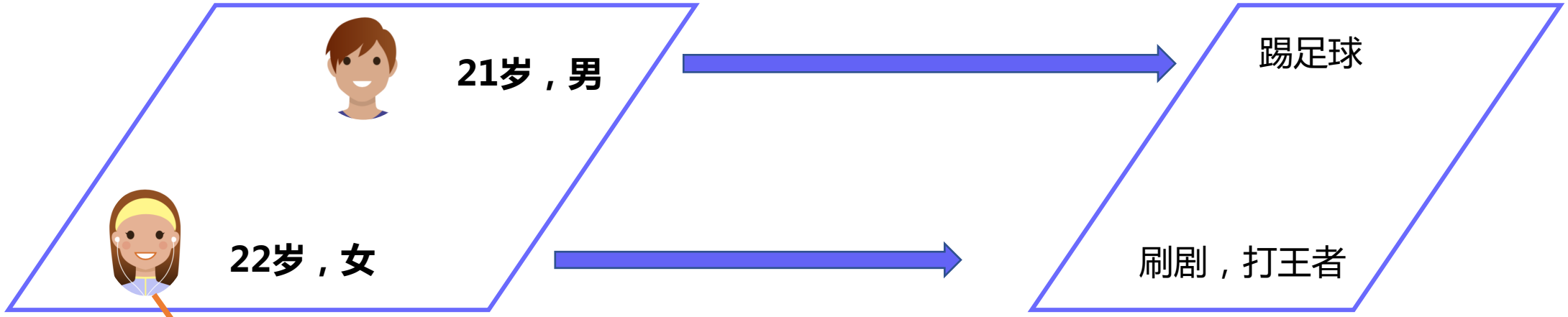
对象与对象之间缺乏关联



提出背景

训练数据

标签



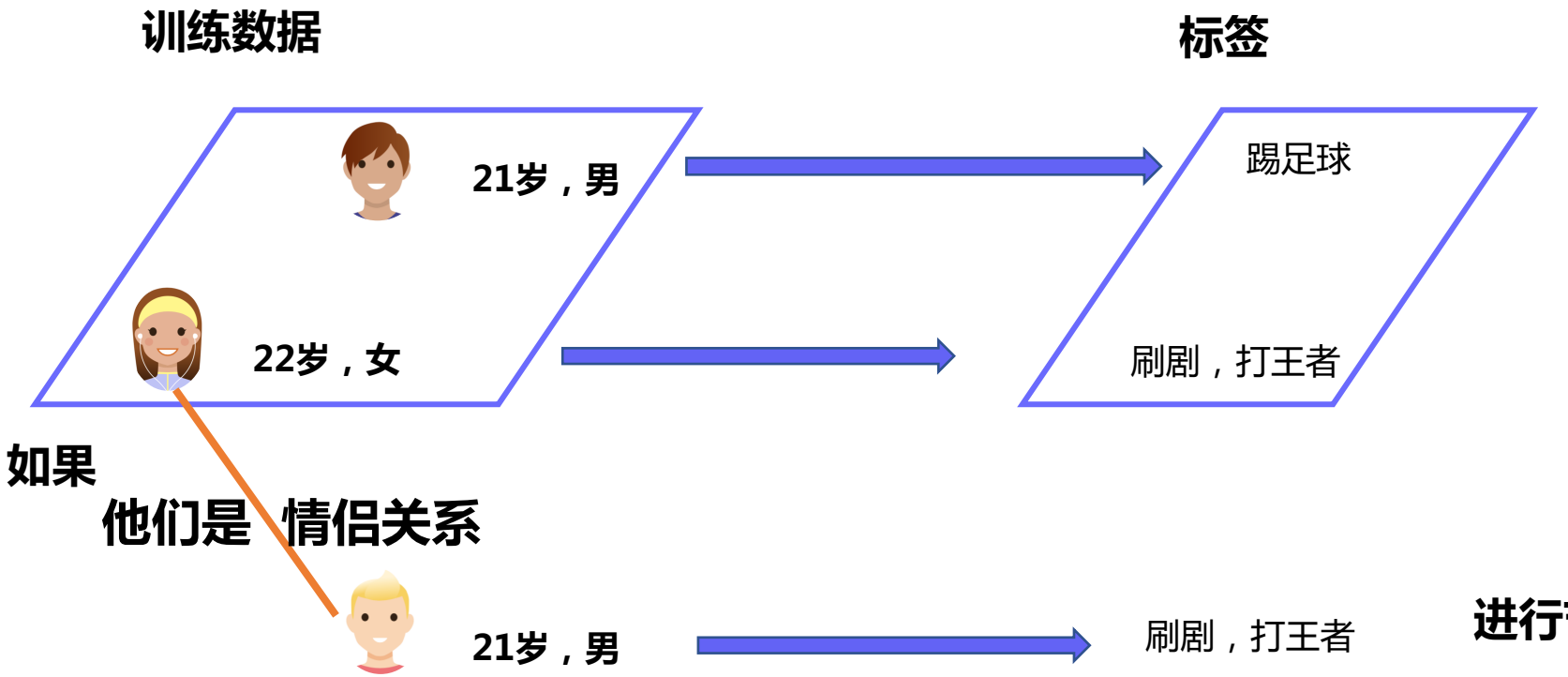
如果

他们是 情侣关系

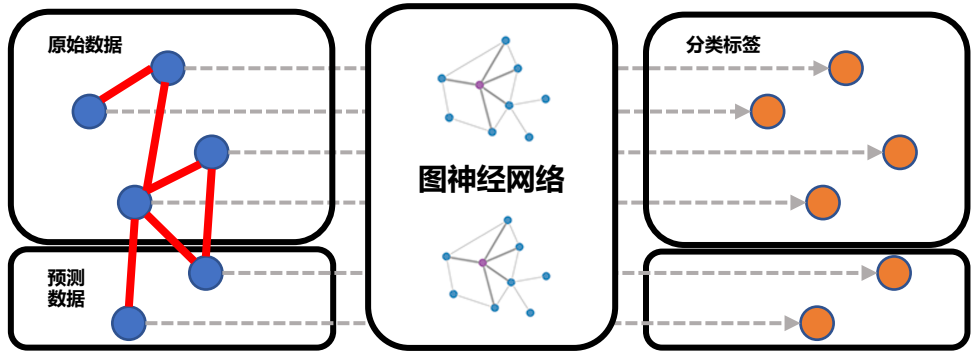
21岁，男

刷剧，打王者

提出背景



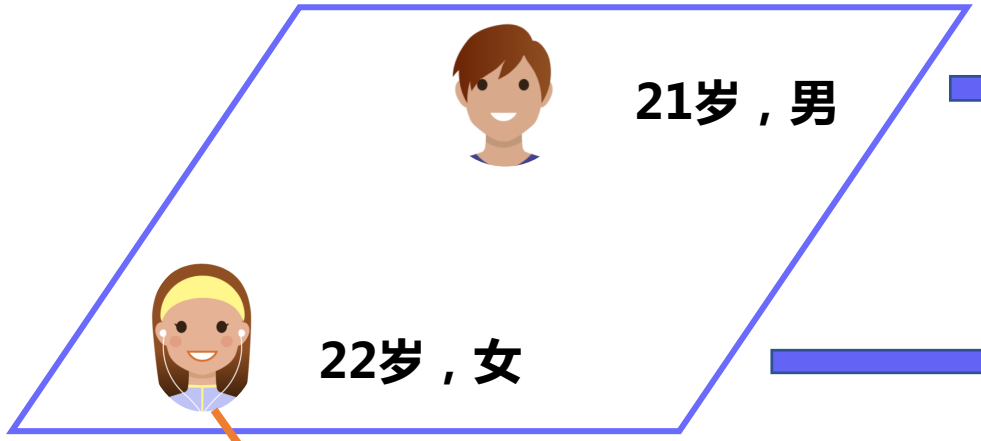
通过图神经网络
进行节点（对象）之间的关系推断

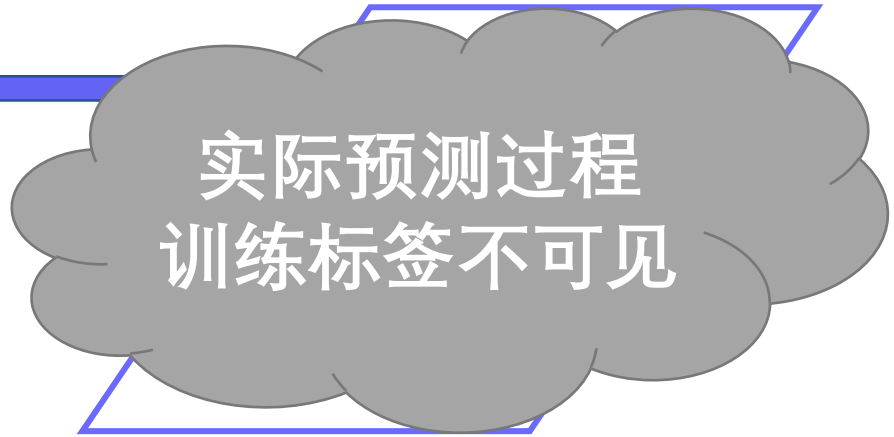


提出背景

训练数据

标签





如果

他们是 情侣关系

提出背景

训练数据

训练集的标签可以作为特征

标签

21岁，男　喜欢：踢足球

22岁，女　喜欢：刷剧，打王者

实际预测过程
训练标签不可见

如果

他们是 情侣关系

标签传播

21岁，男

刷剧，打王者

标签泄露

训练数据

训练集的标签可以作为特征

标签

21岁，男 喜欢：踢足球

标签泄漏

踢足球

22岁，女 喜欢：刷剧，打王者

标签泄漏

刷剧，打王者

21岁，男 喜欢：刷剧，打王者

标签泄漏

刷剧，打王者

Masked Label Prediction

什么是Masked Label Prediction

什么是Masked Label Prediction

训练数据

训练集的标签可以作为特征

标签

21岁，男 喜欢：踢足球 → 踢足球

22岁，女 喜欢：刷剧，打王者 → 刷剧，打王者

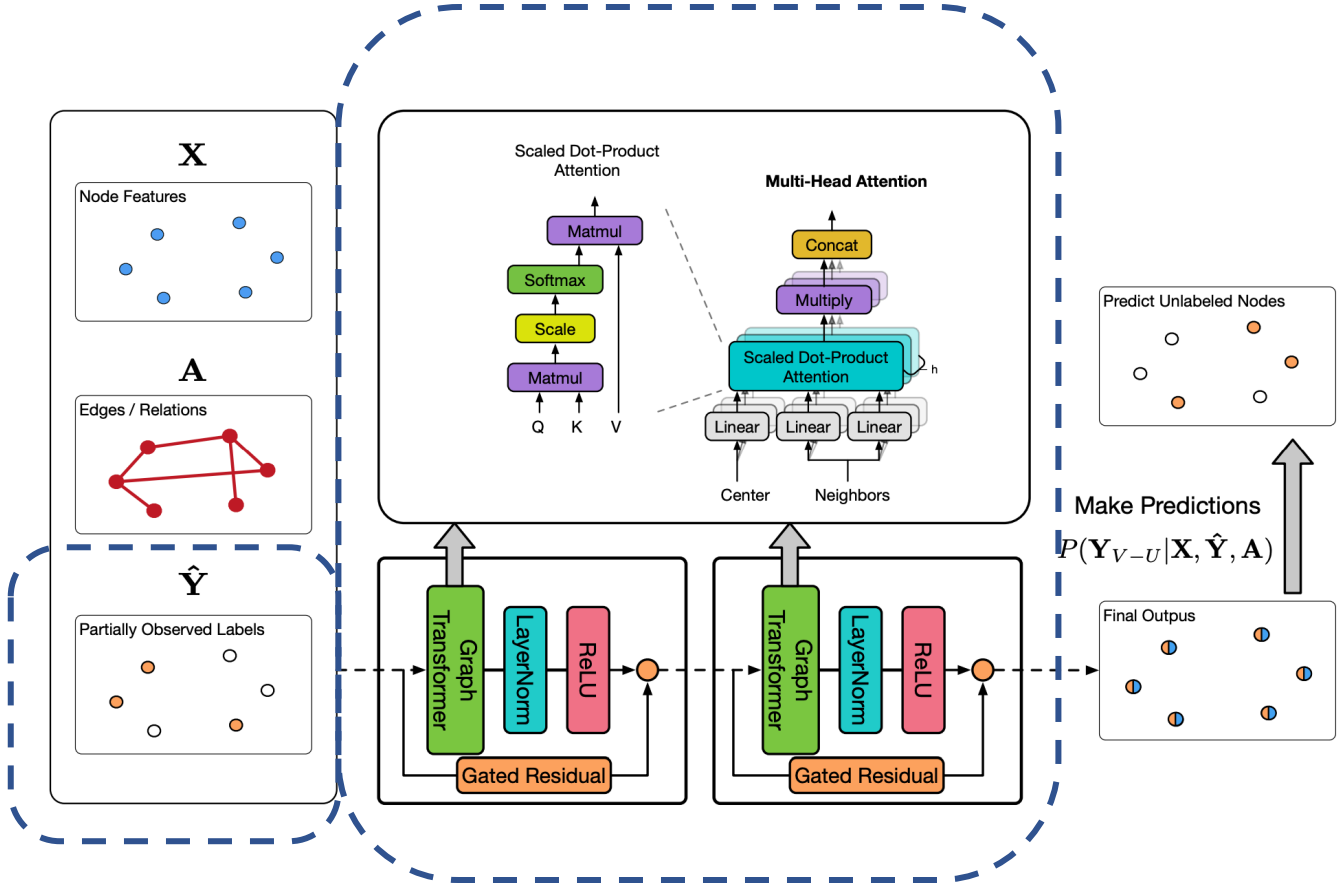
通过Mask预测来学习标签之间的关系

21岁，男 喜欢：刷剧，打王者

Mask → **预测** → 刷剧，打王者

模型结构

使用Graph Transformer替代GAT和GCN



加入训练集标签信息

1. Label Embedding:

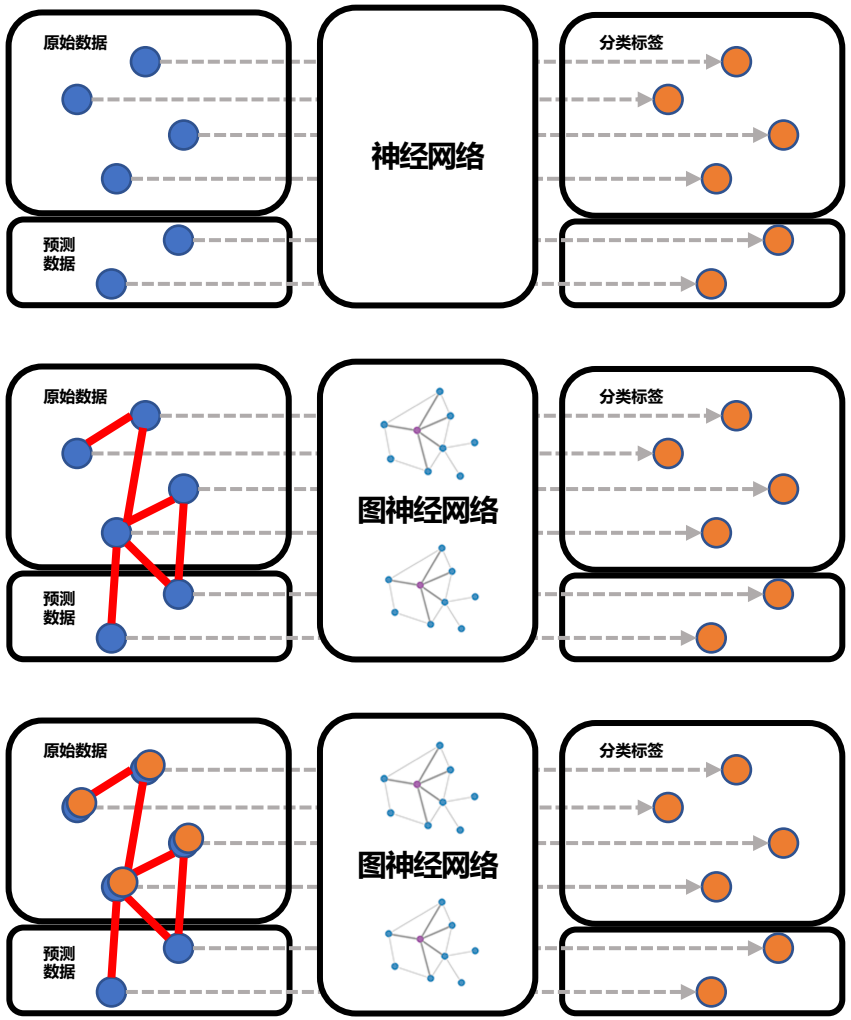
$$\hat{Y} \in \mathbb{R}^{n\times c} \rightarrow \hat{Y}_e \in \mathbb{R}^{n\times f}$$

2. Label Propagation:

$$(H^0 = X + \hat{Y}_e) \in \mathbb{R}^{n\times f}$$

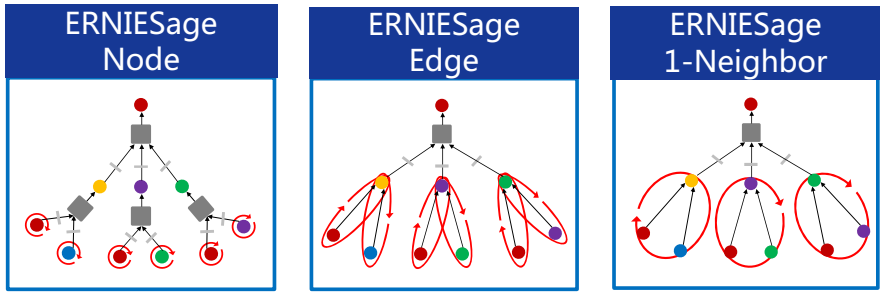
3. 将H^0作为Graph Transformer 输入，参与传播过程

标签传播+图神经网络

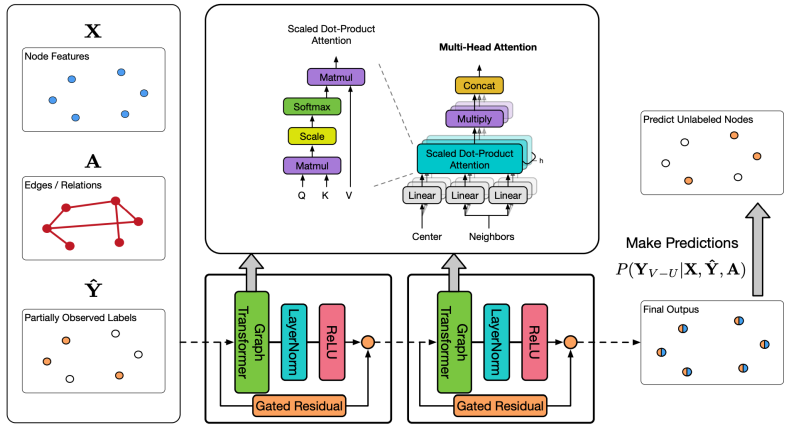


	商品推荐	生物计算	引用网络
单独节点	61%	72%	58%
+节点关系	80%	82%	71%
+标签传播	**83%**	**86%**	**73%**

总结



ERNIESage



UniMP

MASKED LABEL PREDICTION: UNIFIED MESSAGE PASSING MODEL FOR SEMI-SUPERVISED CLASSIFICATION

课堂实践

有请ERNIESage作者讲解模型框架代码

结业比赛

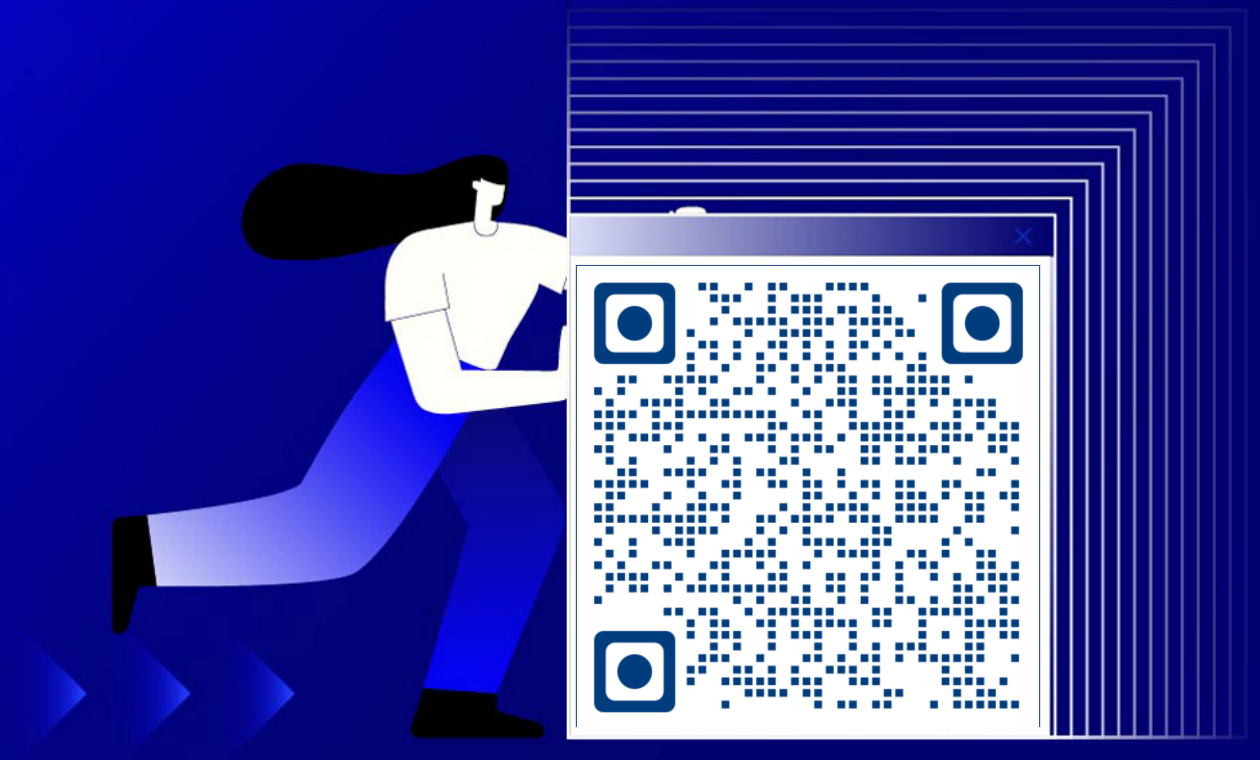
有请比赛发起人讲解

课程大纲

- Graph Learning
- 图学习

参考材料：

- 斯坦福CS224W课程： http://cs224w.stanford.edu
- 图学习库 PGL：https://github.com/PaddlePaddle/PGL



谢谢观看