

Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心



基于图神经网络的中文短文 本匹配方法

硕士研究生 杨俊楠

2020年06月27日

内容提要



- 背景简介
- 基本概念
 - 图神经网络
 - Bert预训练
- 算法原理
- 优劣分析
- 应用总结
- 参考文献

- 预期收获
 - 1. 了解文本匹配的基本概念和研究现状。
 - 2. 了解图神经网络在文本匹配领域的应用方法。

- 文本匹配的概念
 - 文本匹配是比较两个文本并确定它们之间关系的任务。
- 案例一：

蚂蚁金服的NLP比赛-文本语义相似度赛题

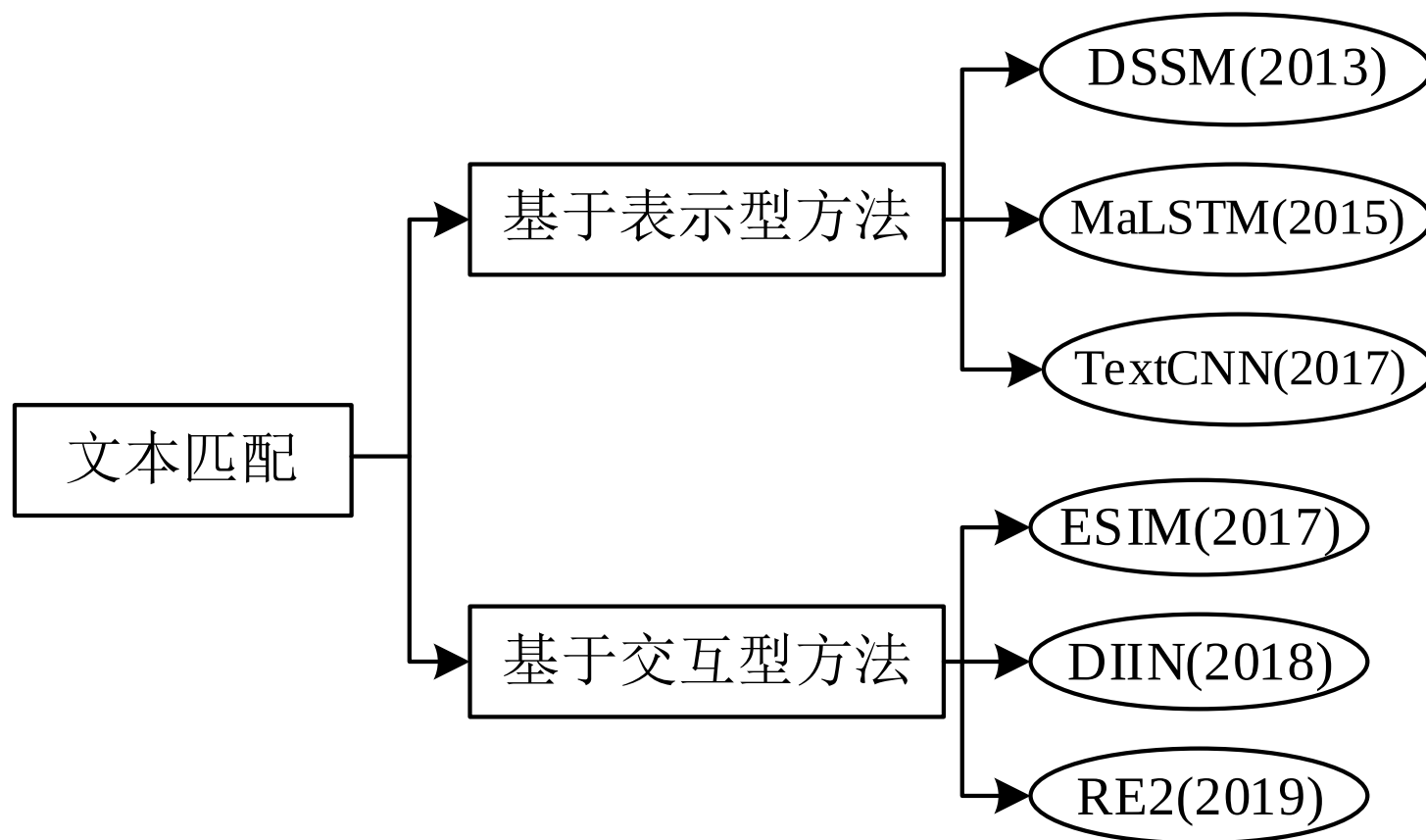
 - 同义问句：“花呗如何还款” – “我怎么还我的花呗呢”
 - 非同义问句：“花呗分期后逾期了如何还款” -- “花呗分期后逾期了哪里还款”

- 案例二

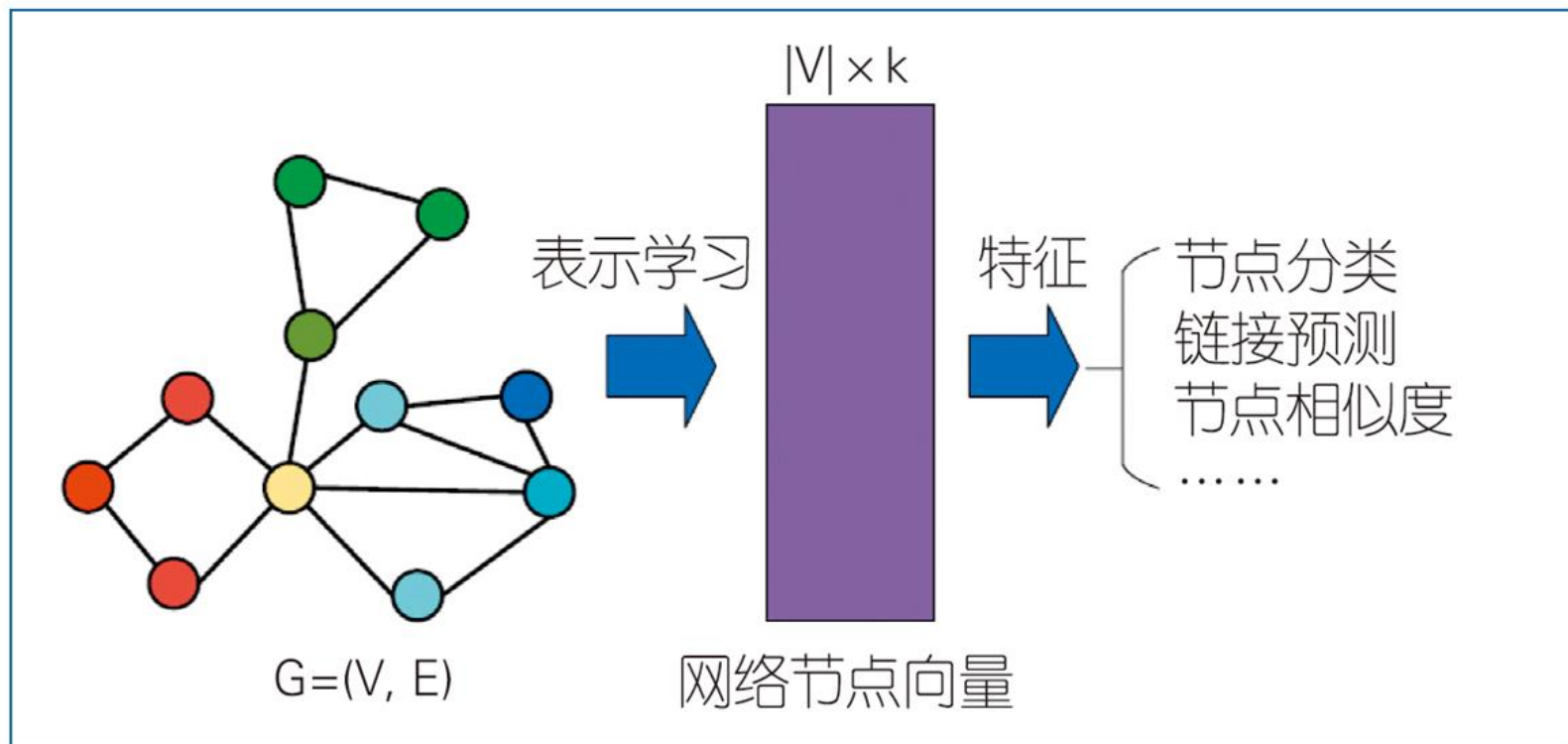
在线广告场景中，需要计算用户查询和页面的语义相似度。



- 研究现状



- 网络嵌入 (Network Embedding)

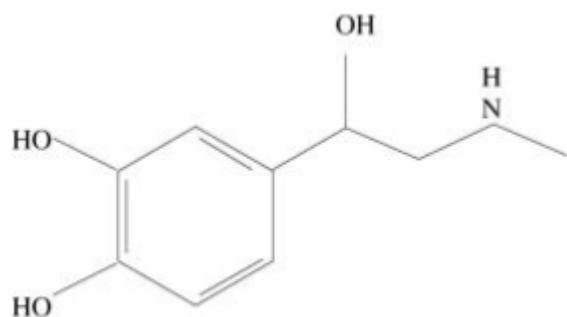


- 图神经网络（GNN）

图神经网络是一种直接作用于图结构上的神经网络。之前的方法都是先将图预处理成能被神经网络处理的格式，例如将图表示成结点序列，但这会丢失很多拓扑信息。**GNN**可以直接处理图，将图或者图中的结点映射为向量。

```
MAIN
  initialize  $w$ ;
   $x = \text{Forward}(w)$ ;
  repeat
     $\frac{\partial \epsilon_w}{\partial w} = \text{BACKWARD}(x, w)$ ;
     $w = w - \lambda \cdot \frac{\partial \epsilon_w}{\partial w}$ ;
     $x = \text{FORWARD}(w)$ ;
  until (a stopping criterion);
  return  $w$ ;
end
```


- 图神经网络 (GNN)



```
FORWARD( $w$ )  
  initialize  $\mathbf{x}(0)$ ,  $t = 0$ ;  
  repeat  
     $\mathbf{x}(t + 1) = F_w(\mathbf{x}(t), l)$ ;  
     $t = t + 1$ ;  
  until  $\|\mathbf{x}(t) - \mathbf{x}(t - 1)\| \leq \varepsilon_f$   
  return  $\mathbf{x}(t)$ ;  
end
```

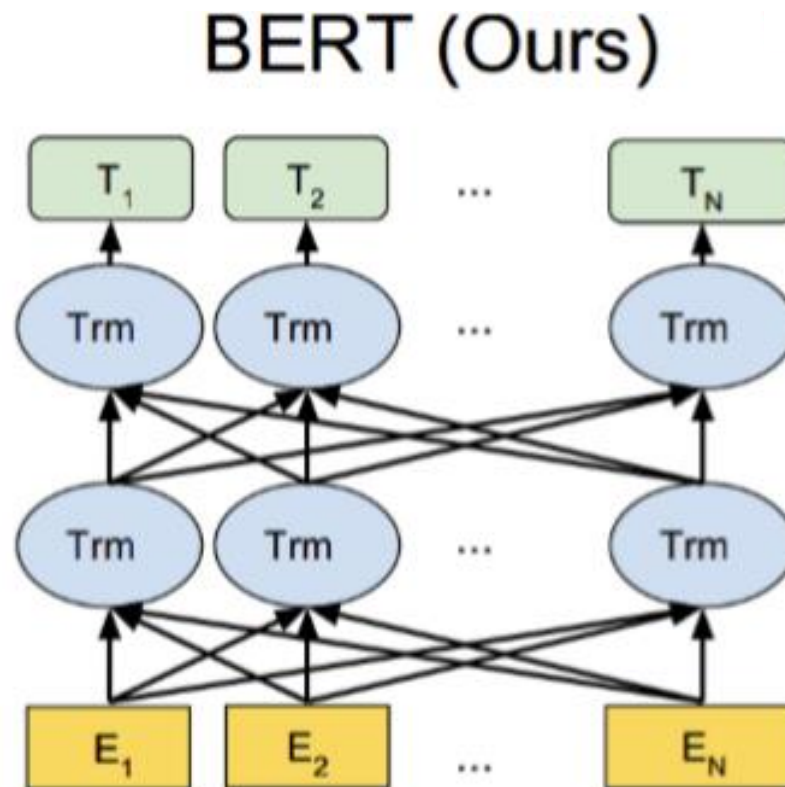
- 图神经网络 (GNN)

```
BACKWARD( $x, w$ )  
   $o = G_w(x, l_N);$   
   $A = \frac{\partial F_w}{\partial x}(x, l);$   
   $b = \frac{\partial e_w}{\partial o} \cdot \frac{\partial G_w}{\partial x}(x, l_N);$   
  initialize  $z(0), t=0;$   
  repeat  
     $z(t) = z(t+1) \cdot A + b;$   
     $t=t-1;$   
  until  $\|z(t-1) - z(t)\| \leq \varepsilon_b;$   
   $c = \frac{\partial e_w}{\partial o} \cdot \frac{\partial G_w}{\partial w}(x, l_N);$   
   $d = z(t) \cdot \frac{\partial F_w}{\partial w}(x, l);$   
   $\frac{\partial e_w}{\partial w} = c + d;$   
  return  $\frac{\partial e_w}{\partial w};$   
end
```

- Bert预训练

Bert是双层双向Transformer结构，采用两阶段模型，首先是语言模型预训练；其次是使用Fine-Tuning模式解决下游任务。

Bert对句子关系类下游任务改造：加上一个起始和终结符号，句子之间加分隔符。



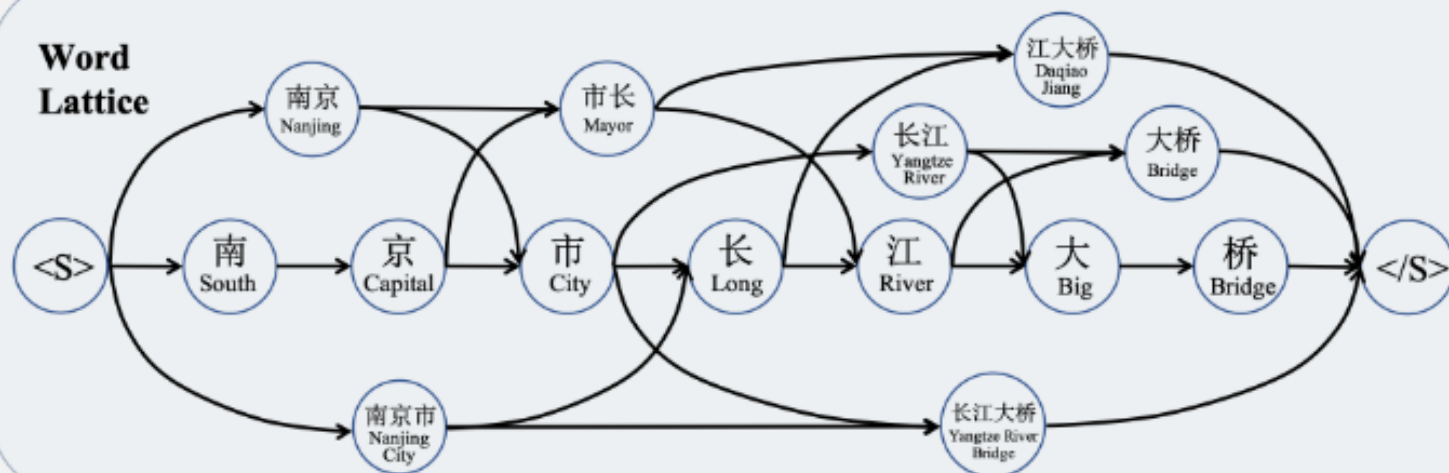
- 基于图神经网络的中文短文本匹配
 - 中文文本在匹配时面临分词错误、不一致、歧义等问题。

Original Sentence: 南京市长江大桥 (South | Capital | City | Long | River | Big | Bridge)

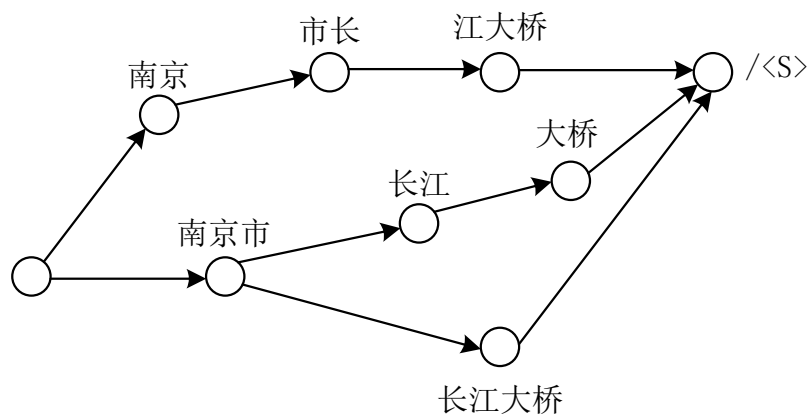
Segment-1: 南京市 / 长江大桥 (Nanjing City / Yangtze River Bridge)

Segment-2: 南京市 / 长江 / 大桥 (Nanjing City / Yangtze River / Bridge)

Segment-3: 南京 / 市长 / 江大桥 (Nanjing / Mayor / Daqiao Jiang)

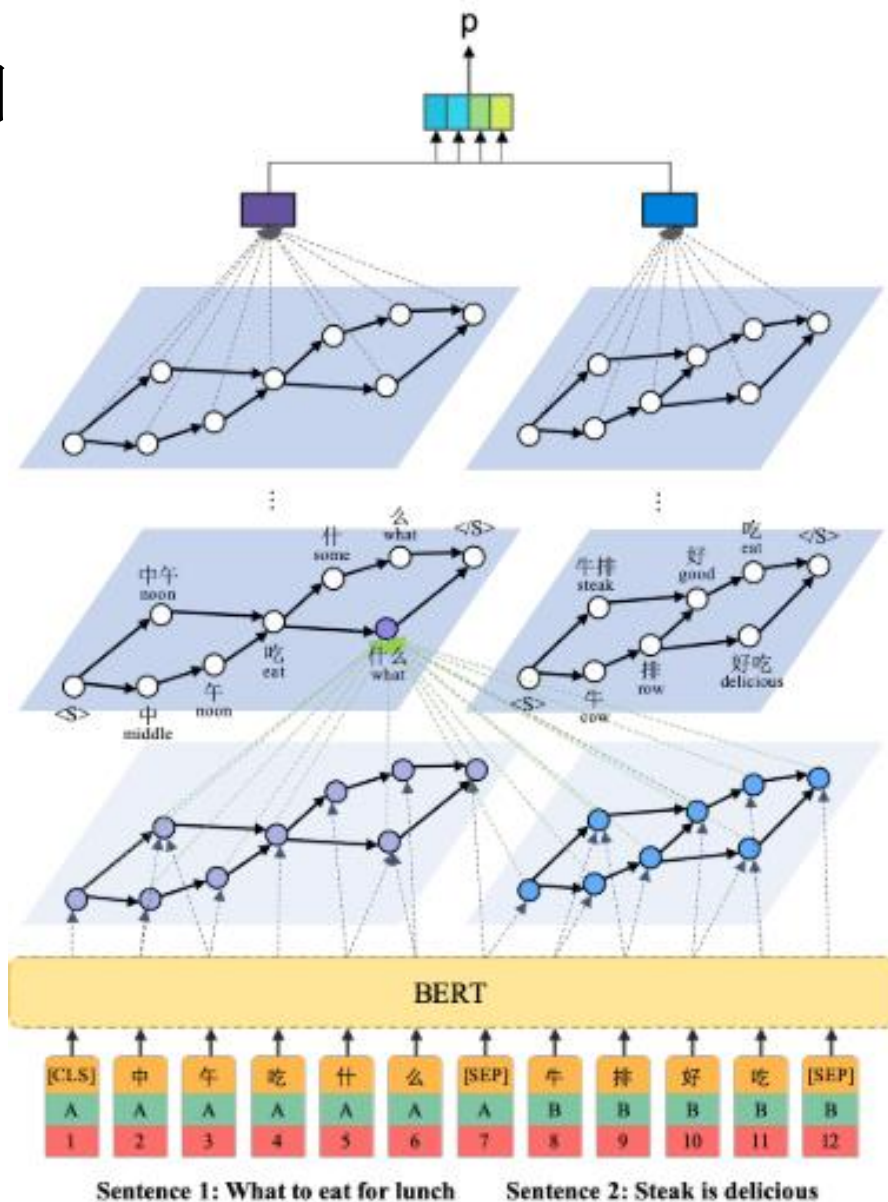


- 如何构建图？
- 不同分词序列组合为词格（Lattice）



T	提高中文短文本匹配的性能
I	中文短文本对
P	Bert预训练词向量作为节点特征向量 For step= 1, L do: 1. 节点信息传播 2. 文本相似度计算 3. 节点状态更新 文本匹配结果预测
O	匹配结果
P	解决中文分词错误、歧义、不一致导致匹配效果差问题
C	图神经网络有强大的表示学习能力
D	如何将文本数据应用于图神经网络
L	ACL2020

- 算法原理图



匹配结果预测层

图神经网络匹配层

上下文节点嵌入层

- 上下文节点嵌入
 - Bert预训练字向量

$$\{[\text{CLS}], c_1^a, \dots, c_{t_a}^a, [\text{SEP}], c_1^b, \dots, c_{t_b}^b, [\text{SEP}]\}$$

- 字向量加权计算词向量, 假设一个词由 n_i 个字组成 $\{c_{s_i}, c_{s_i+1}, \dots, c_{s_i+n_i-1}\}$

$$\begin{aligned}\hat{\mathbf{u}}_{s_i+k} &= \text{FFN}(\mathbf{c}_{s_i+k}) \\ \mathbf{v}_i &= \sum_{k=0}^{n-1} \mathbf{u}_{s_i+k} \odot \mathbf{c}_{s_i+k}\end{aligned}$$

由此计算的词向量作为节点的初始向量。

- 图神经网络匹配

- 信息传播

$$\begin{aligned}\mathbf{m}_i^{fw} &= \sum_{v_j \in \mathcal{N}_{fw}(v_i)} \alpha_{ij} \left(\mathbf{W}^{fw} \mathbf{h}_j^{l-1} \right) & \mathbf{m}_i^{b1} &= \sum_{v_m \in \mathcal{V}^b} \alpha_{im} \left(\mathbf{W}^{fw} \mathbf{h}_m^{l-1} \right) \\ \mathbf{m}_i^{bw} &= \sum_{v_k \in \mathcal{N}_{bw}(v_i)} \alpha_{ik} \left(\mathbf{W}^{bw} \mathbf{h}_k^{l-1} \right) & \mathbf{m}_i^{b2} &= \sum_{v_q \in \mathcal{V}^b} \alpha_{iq} \left(\mathbf{W}^{bw} \mathbf{h}_q^{l-1} \right)\end{aligned}$$

定义: $\mathbf{m}_i^{self} \triangleq [\mathbf{m}_i^{fw}, \mathbf{m}_i^{bw}]$, $\mathbf{m}_i^{cross} \triangleq [\mathbf{m}_i^{b1}, \mathbf{m}_i^{b2}]$

- 节点更新:

$$d_k = \text{cosine} \left(\mathbf{w}_k^{cos} \odot \mathbf{m}_i^{self}, \mathbf{w}_k^{cos} \odot \mathbf{m}_i^{cross} \right)$$

$$\mathbf{h}_i^l = \text{FFN} \left(\left[\mathbf{m}_i^{self}, \mathbf{d}_i \right] \right)$$

其中: $k \in \{1, 2, \dots, P\}$, $\mathbf{d}_i \triangleq [d_1, d_2, \dots, d_P]$.

- 匹配结果预测
 - 关系预测
 - 图上节点向量进行attentive pooling得到图向量表示

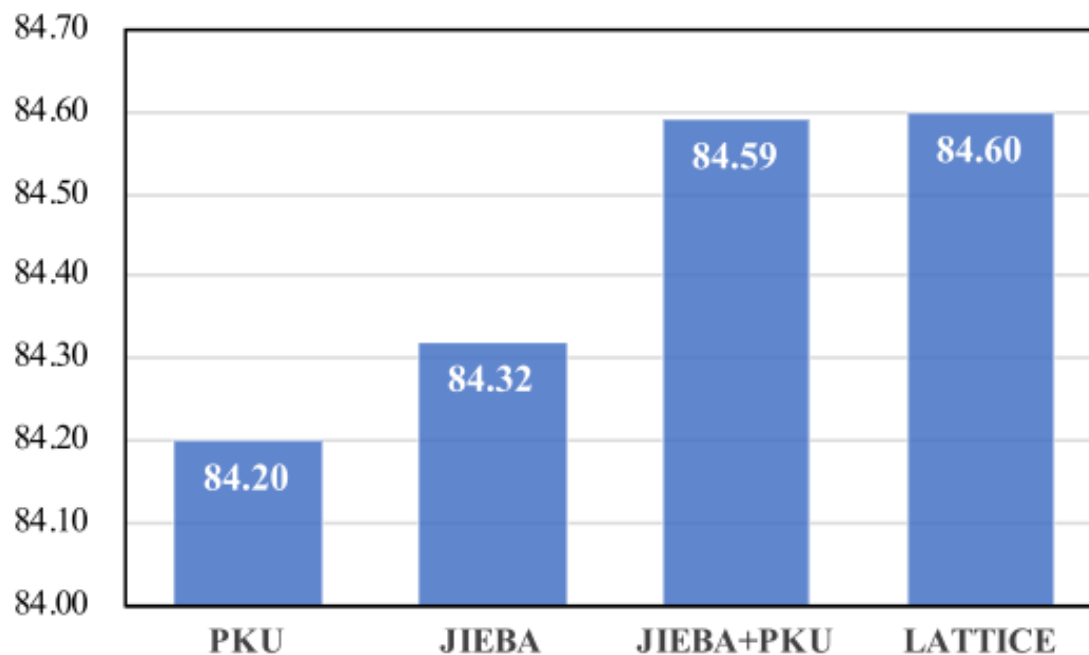
$$p = \text{FFN} \left(\left[\mathbf{g}^a, \mathbf{g}^b, \mathbf{g}^a \odot \mathbf{g}^b, |\mathbf{g}^a - \mathbf{g}^b| \right] \right)$$

- 实验结果

Models	BQ		LCQMC	
	ACC.	F1	ACC.	F1
Text-CNN	68.5	69.2	72.8	75.7
BiLSTM	73.5	72.7	76.1	78.9
Lattice-CNN	78.2	78.3	82.1	82.4
BiMPM	81.9	81.7	83.3	84.9
ESIM-char	79.2	79.3	82.0	84.0
ESIM-word	81.9	81.9	82.6	84.5
GMN (Ours)	84.2	84.1	84.6	86.0
BERT	84.5	84.0	85.7	86.8
BERT-wwm	84.9	-	86.8	-
BERT-wwm-ext	84.8	-	86.6	-
ERNIE	84.6	-	87.0	-
GMN-BERT (Ours)	85.6	85.5	87.3	88.0

预训练过的模型表现优于没有预训练的模型，基于交互型的方法优于基于表示型的方法

- 实验结果



采用多粒度分词路径作为输入，效果优于只输入一种分词序列

- 自然语言推理
- 文本释义
- 信息检索
- 问答系统
- 推荐系统
- ...

- Scarselli F , Gori M , Tsoi A C , et al. The Graph Neural Network Model[J]. IEEE Transactions on Neural Networks, 2009, 20(1):61-80.
- <https://zhuanlan.zhihu.com/p/76290138>
- <https://zhuanlan.zhihu.com/p/85768094>
- https://blog.csdn.net/weixin_43269174/article/details/98492487



谢谢！

大成若缺，其用不弊。大盈
若冲，其用不穷。大直若屈。
大巧若拙。大辩若讷。静胜
躁，寒胜热。清静为天下正。

