



大连理工大学

信息检索研究室

Information retrieval laboratory of DUT



# 临床试验筛选标准 短文本分类 CHIP2019-评测任务3

参赛队: DUTIR914

大连理工大学-信息检索研究室

汇报人: 李孟颖

2019年11月

1

任务介绍

2

实验方案

3

实验结果

4

总结展望



大连理工大学

信息检索研究室

Information retrieval laboratory of DLUT



# 任务介绍

任务：判断给定的一系列中文临床医学描述句子的类别

44类

ID	输入(筛选标准)	输出(类别)
S1	年龄>80岁	Age
S2	近期颅内或椎管内手术史	Therapy or Surgery
S3	血糖<2.7mmol/L	Laboratory Examinations

## ● 数据分布

◆ 训练集：验证集：测试集=3:1:1

◆ 无监督数据集：213,154

◆ 训练集：22,962

◆ 验证集：7,682

◆ 测试集：7,697

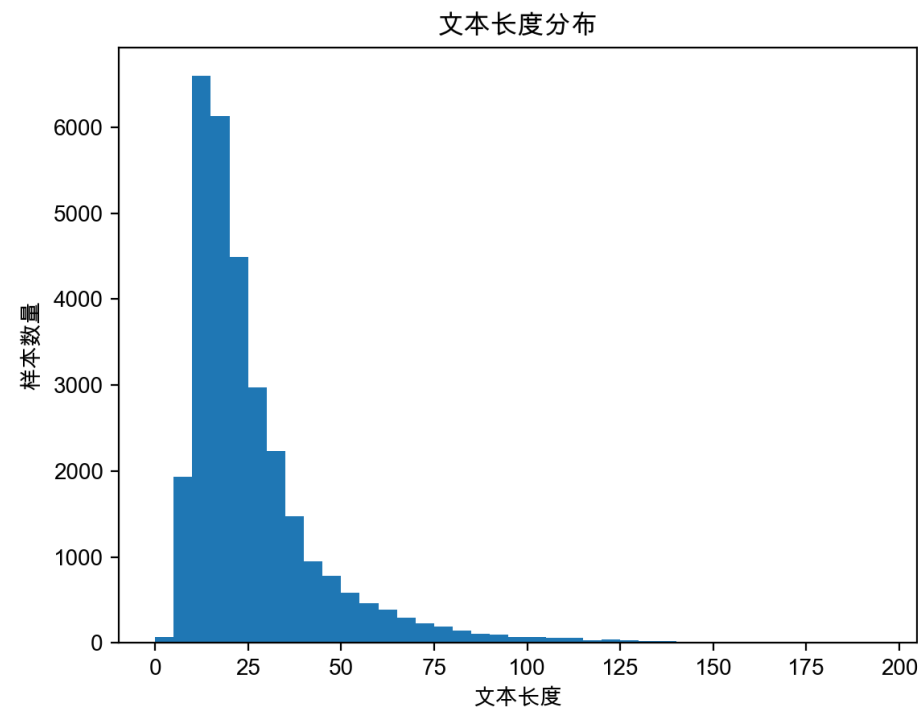
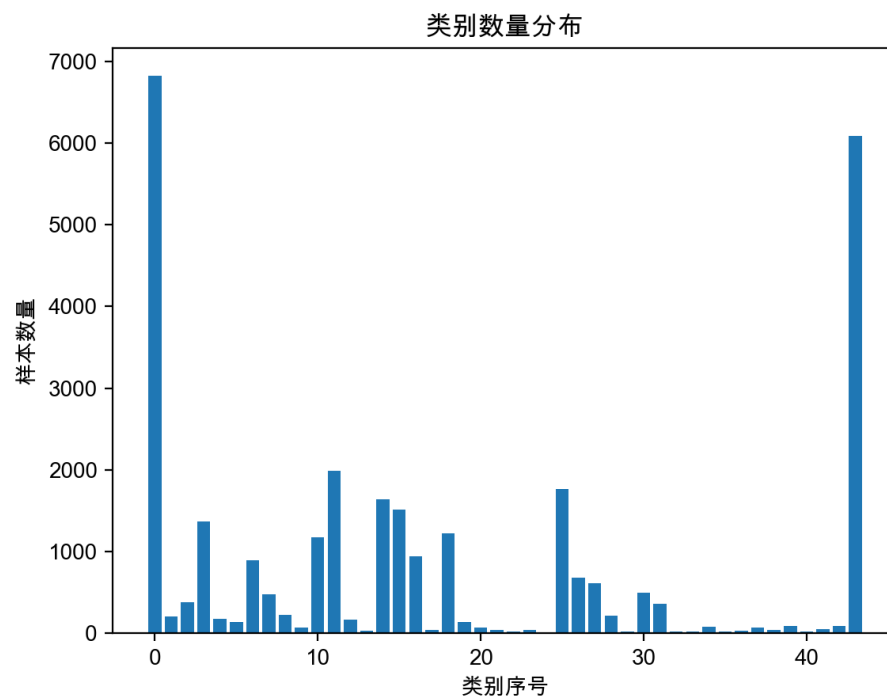
## ● 数据分析

◆ 数据样本不平衡

## ● 句子长度分布

◆ 最长：341

◆ 最短：2





# 实验方案

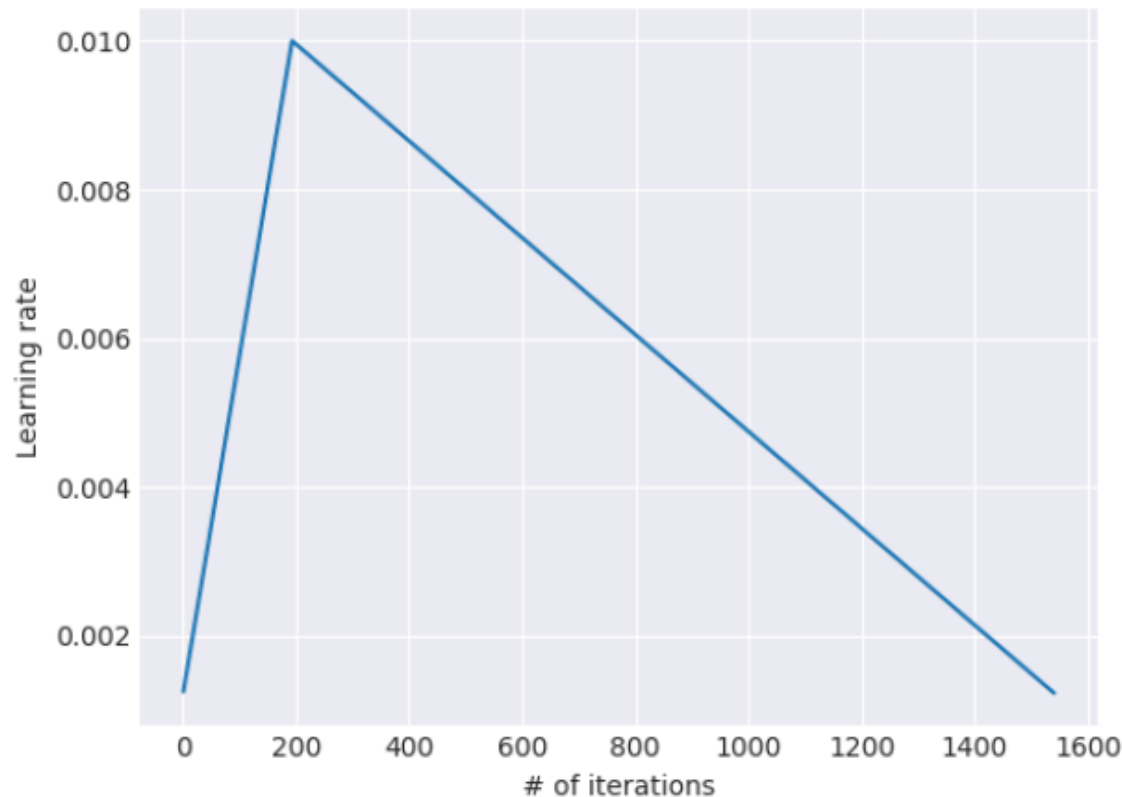
- **三个阶段**
- 使用预训练语言模型，并对语言模型进行微调
- 将语言模型与神经网络模型结合得到分类模型，并对分类模型进行微调
- 将多个分类模型进行集成学习得到最终的模型分类器

## ● LM

- 哈工大讯飞联合实验室发布的RoBERTa-wwm-ext, Chinese
- Google发布的中文版 BERT-base, Chinese  
(<https://github.com/ymcui/Chinese-BERT-wwm>)

## ● fine-tuning

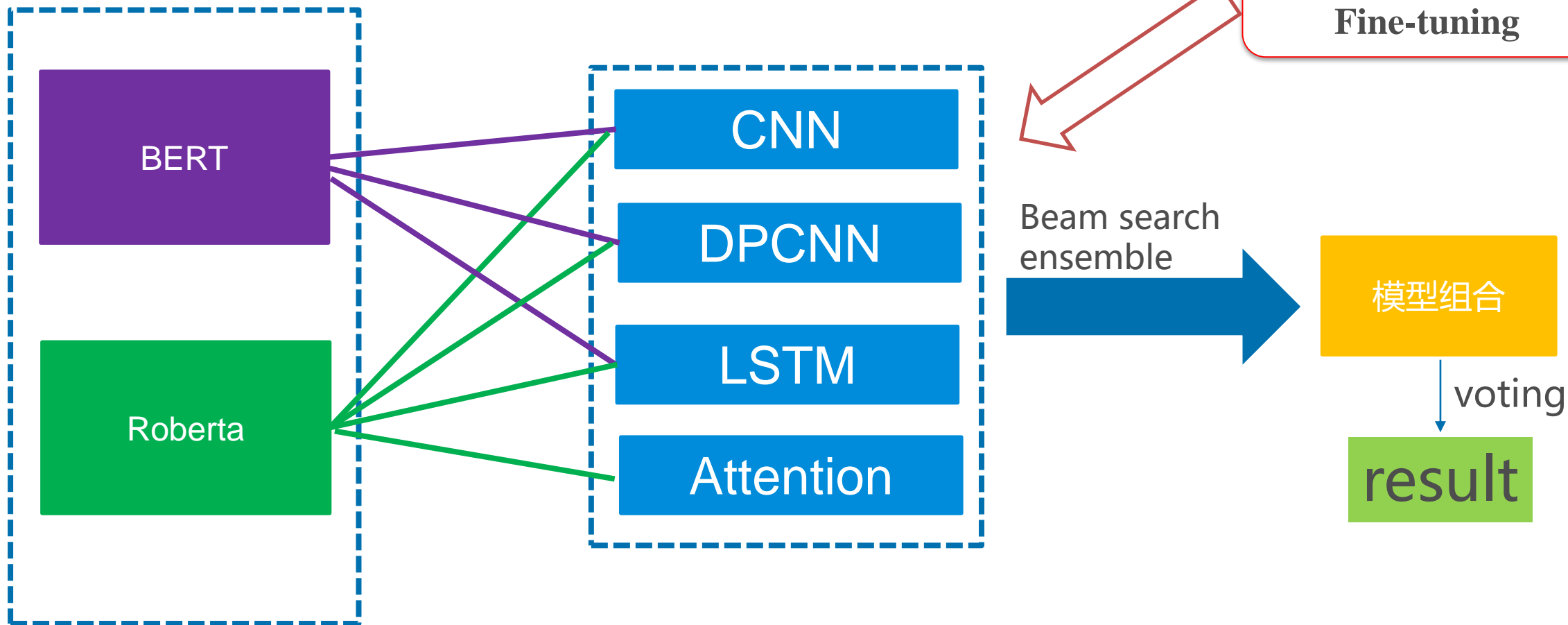
- 使用213154无监督数据
- 用发布的训练集和验证集做结果验证
- 使用倾斜的三角学习率方法



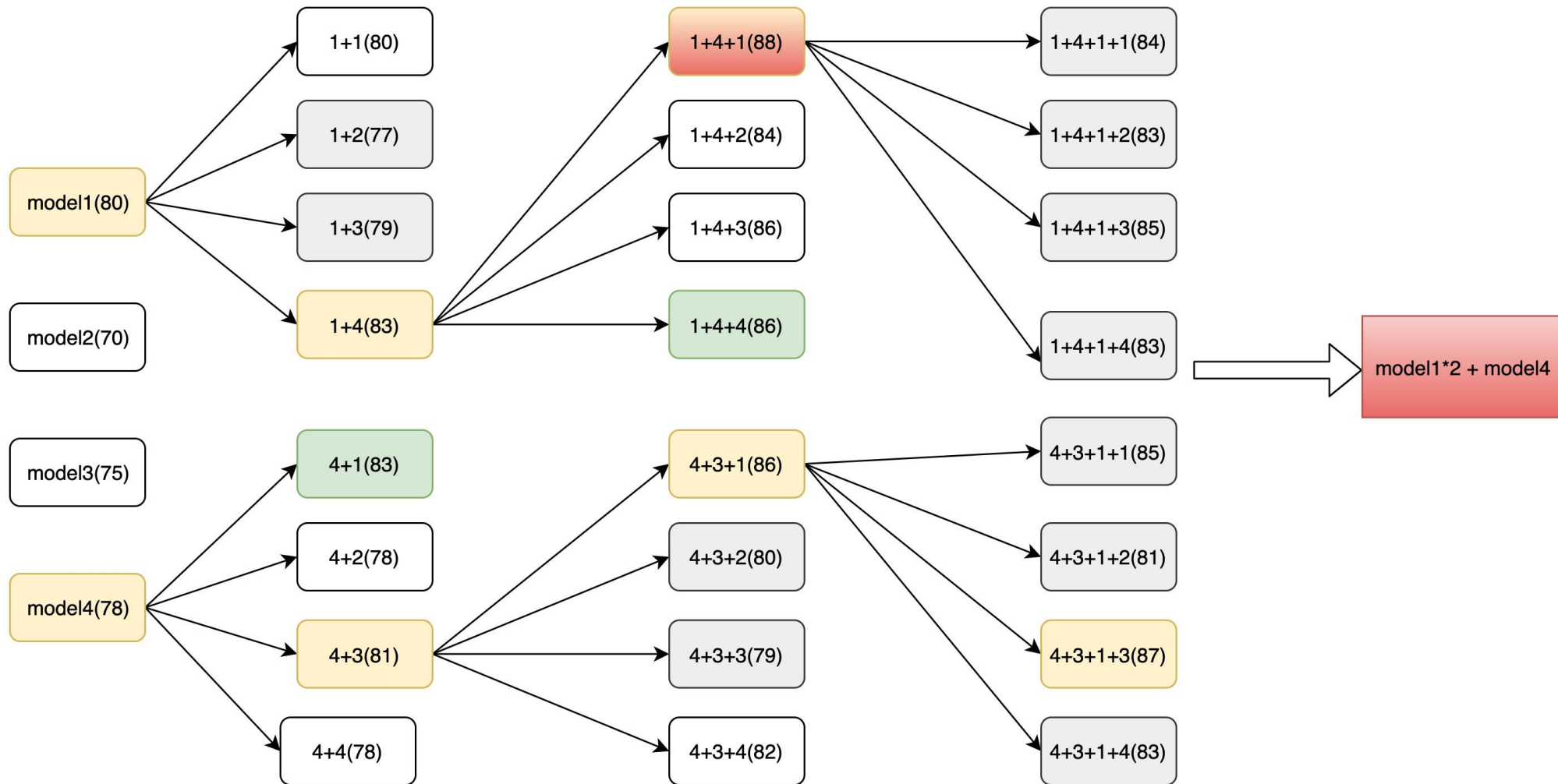


预训练模型 (fine-tuning)

神经网络模型



## Beam search ensemble算法



- Focal loss函数
- 为解决样本不平衡问题
- $FL_{(P_t)} = -\alpha_t (1 - P_t)^\gamma \log (P_t)$
- 其中,  $\alpha_t$ 表示每个类别的权重系数,  $C_t$ 表示训练集中第t个类别的个数,  $\gamma$ 为调制系数, 实验时设置 $\gamma=2$
- 当 $\gamma=0$ 的时候, focal loss就是传统的交叉熵损失, 当 $\gamma$ 增加的时候, 调制系数也会增加。其核心思想是用一个合适的函数去度量难分类和易分类样本对总的损失的贡献, 这样既能调整正负样本的权重, 又能控制难易分类样本的权重。

Lin T Y , Goyal P , Girshick R , et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.



# 实验结果

# 最终结果



大连理工大学

信息检索研究室

Information retrieval laboratory of DUT



方法	得分
bert_base	0.7964
bert_dpcnn	0.7976
bert_lstm_att	0.8049
bert_cnn	0.8012
roberta_base	0.7990
roberta_att	0.7948
roberta_dpcnn	0.8030
roberta_lstm_att	0.8028
roberta_cnn	0.8063
all_voting	<b>0.8099</b>
beam_search_voting	<b>0.8111</b>

#Rank

1  
2  
3  
4  
5

队伍名

wzm  
DUTIR914  
DUTIRTM  
好果汁队  
努力! 奋斗!

分数

0.810263  
0.809936  
0.807456  
0.800168  
0.800123



# 总结展望

- 针对临床试验筛选标准短文本分类任务的实现主要有以下三个部分：

## 语言模型

- 使用bert, roberta两个预训练语言模型
- 使用相关领域的外部资源数据集进行fine-tuning, 使用了倾斜三角学习率的微调方法。

## 分类模型

- 使用CNN, DPCNN, LSTM, Attention等神经网络模型与语言模型结合形成医学文本分类模型。
- 对分类模型进行fine-tuning, 使用了逐层解冻的微调方法。

## 集成学习

- 将上述得到的9个分类模型全部直接进行投票得到一个分类结果。
- 使用beam search ensemble算法得到模型组合进行投票得到一个分类结果。
- 最佳模型组合为bert\_bert+dpcnn, bert+lstm+attention, roberta+attention, roberta+cnn \* 2, roberta+dpcnn, roberta+lstm+attention。

## ● 实验过程

- ◆ 尝试对数据进行去噪处理，结果没有提升。
- ◆ 尝试使用多种预训练语言模型，包括Elmo, Ernie, Xlnet等，但效果均不及bert和roberta，选择两个预训练模型的原因是希望两个预训练模型发挥不同效用学习到不同特征，但是实验结果表明两者学习到的特征相似，实验效果提升不明显。
- ◆ 尝试使用基于机器学习的一些分类方法进行集成，效果均不及单个bert模型。

## ● 优劣势

- ◆ Beam search ensemble算法可以快速找出相对较优的模型组合，其投票的得分决定了最后结果，其中性能相对较好的模型可重复。加入更多模型集成并无提升。
- ◆ 没有加入人工特征，对临床试验筛选标准短文本分类任务的特征分析不足，分类结果有待提升，但也因此节省了时间和精力。



## ● 展望:

- ◆ 神经网络模型的优化和参数调优。
- ◆ 加入原始文本的语言特征，包括词性特征，句法特征等。
- ◆ 针对医学短文本的特征抽取以及选择。
- ◆ 探索生物医学领域相关的一些外部知识并引入到深度模型中。

## ● 致谢:

- ◆ 感谢主办方。
- ◆ 感谢本次比赛的运营人员。
- ◆ 感谢指导老师和队友。



# 谢谢!