

Semantic similarity model of medical questions based on fusion pre-training model

伏冠宇

2019 / 11 / 24

简介

- 平安医疗科技疾病问答迁移学习比赛

- 问句来自互联网上患者真实的问题，并经过了筛选和人工的意图匹配标注
- 问句由5个病种构成

输入：

疾病问句对：“糖尿病吃什么？”，“糖尿病的食谱？”

病种名称：diabetes

输出：

事件主体：1（语义相同）

挑战

■ 问句来自来自互联网上患者的真实问题

- 训练集中不同病重的问句数量不均衡
- 文本噪声：描述不规范；隐晦词的不同表述方法
- 标注噪声：语义模糊样本

■ 病种类型会对问句对语义产生影响

- 某些问句会讨论多个病重

动机

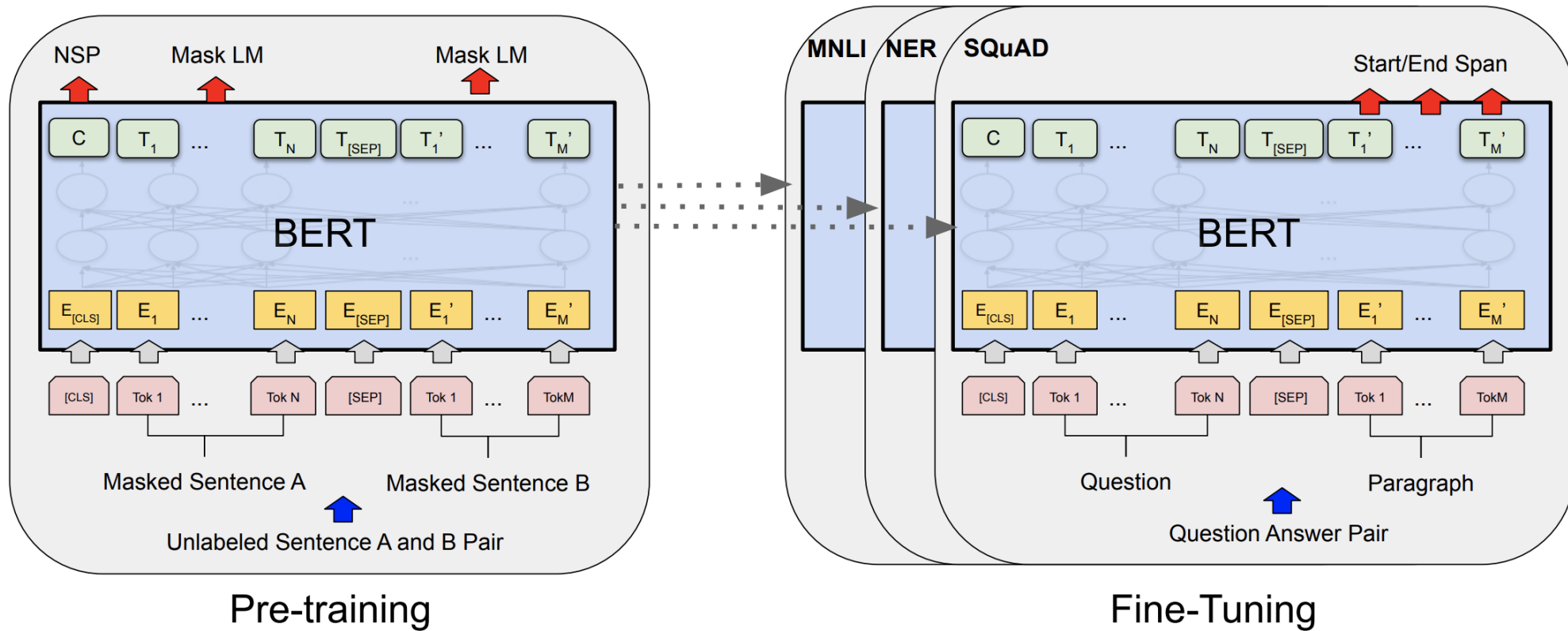
当前预训练模型 Pretrained + Fine-tuning的范式提升了很多 NLP任务的性能

不同预训练模型的性能表现有较大的差距

尝试引入疾病类别的信息

背景

- 预训练模型



背景

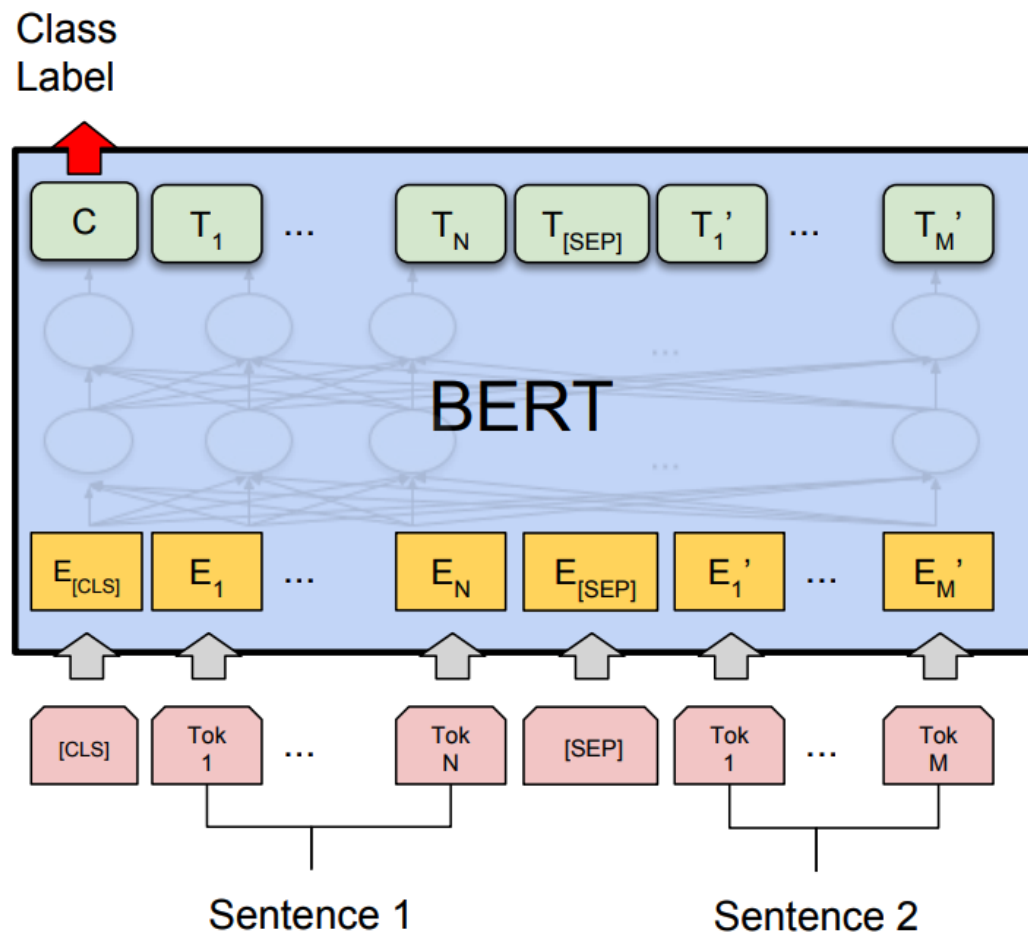
- 时间线

- **ELMo**: Deep Contextualized Word Representations, 2018
- **GPT**: Improving Language Understanding by Generative Pre-Training, 2018
- **BERT**: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018
- **MT-DNN**: Multi-Task Deep Neural Networks for Natural Language Understanding, 2018
- **ERNIE**: Enhanced Language Representation with Informative Entities, 2019
- **XLNET**: Generalized Autoregressive Pretraining, 2019
- **RoBERTa**: Robustly Optimized BERT Pretraining Approach, 2019
-

背景

- BERT

- 多层双向Trans
- 输入为: token
- 两个预训练任务
- 在BooksCorpus
- 适用于多个下游任务
- 生成 ...



tion
(10M words)上预训练
、文本匹配, 文本

背景

- BERT

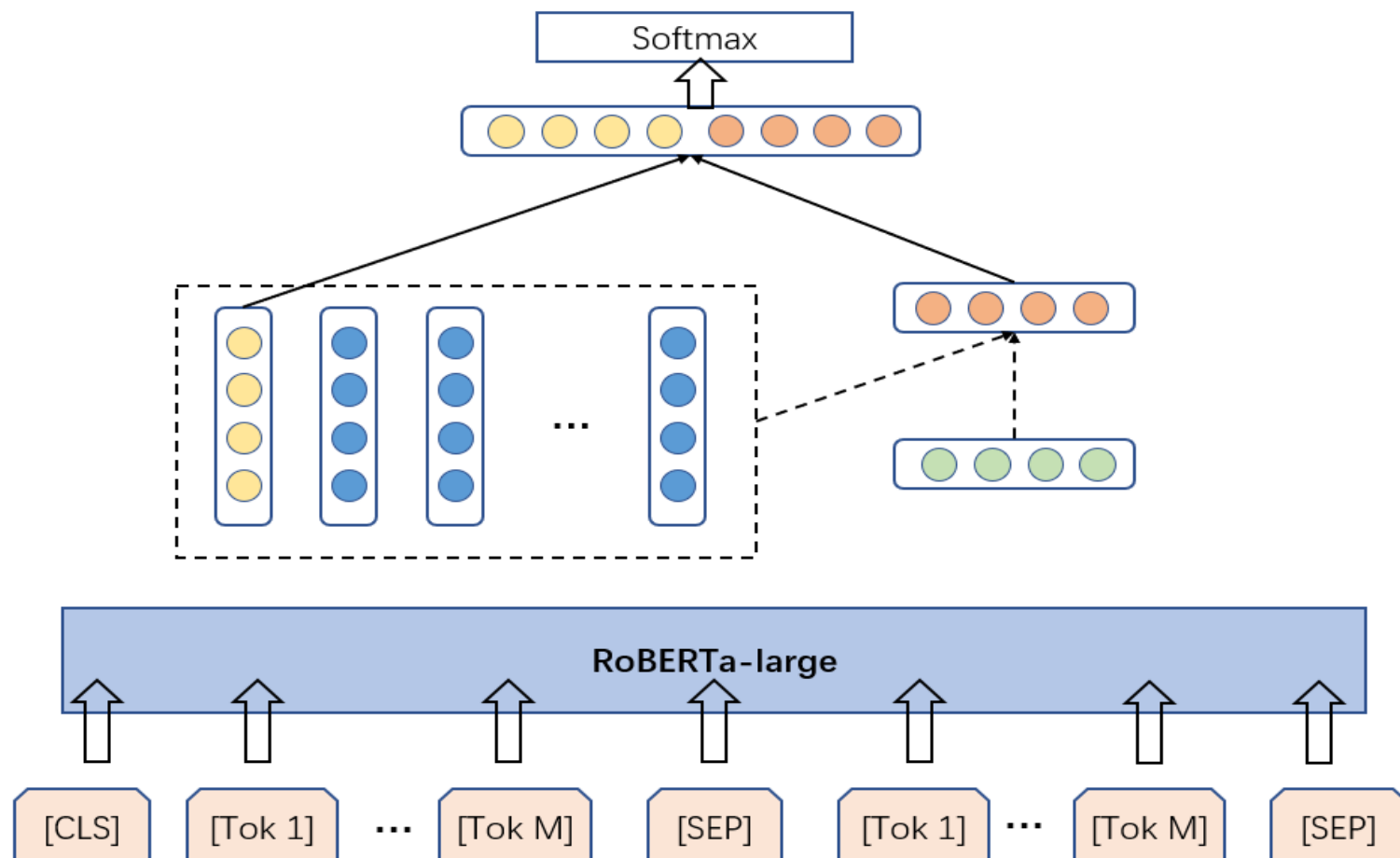
- 刷新11项NLP测试任务，代码和预训练模型公开

- 中文领域:

- 来源于社区贡献，大多数采用中文维基百科和通用数据，以及私有数据预训练

- BERT-base-Chinese: <https://github.com/google-research/bert>
 - ERNIE: <https://github.com/PaddlePaddle/ERNIE/tree/develop/ERNIE>
 - Chinese-BERT-WWM: [Pre-Training with Whole Word Masking for Chinese BERT](#)
 - Chinese-PreTrained-XLNet: <https://github.com/ymcui/Chinese-PreTrained-XLNet>

方法



- 我们在BERT后又加上了一层self-attention编码层，使模型学习到更多的面向任务本身的编码信息
- 构造预训练任务
- 其他尝试
 - R-NET
 - Multi-task
 - 两阶段fine-tuning
 - ...

实验

■ 数据集

- 发布训练数据20000对
- 验证数据有10000对
- 测试数据有5万对，其中只有部分数据参与最终排名

■ 指标

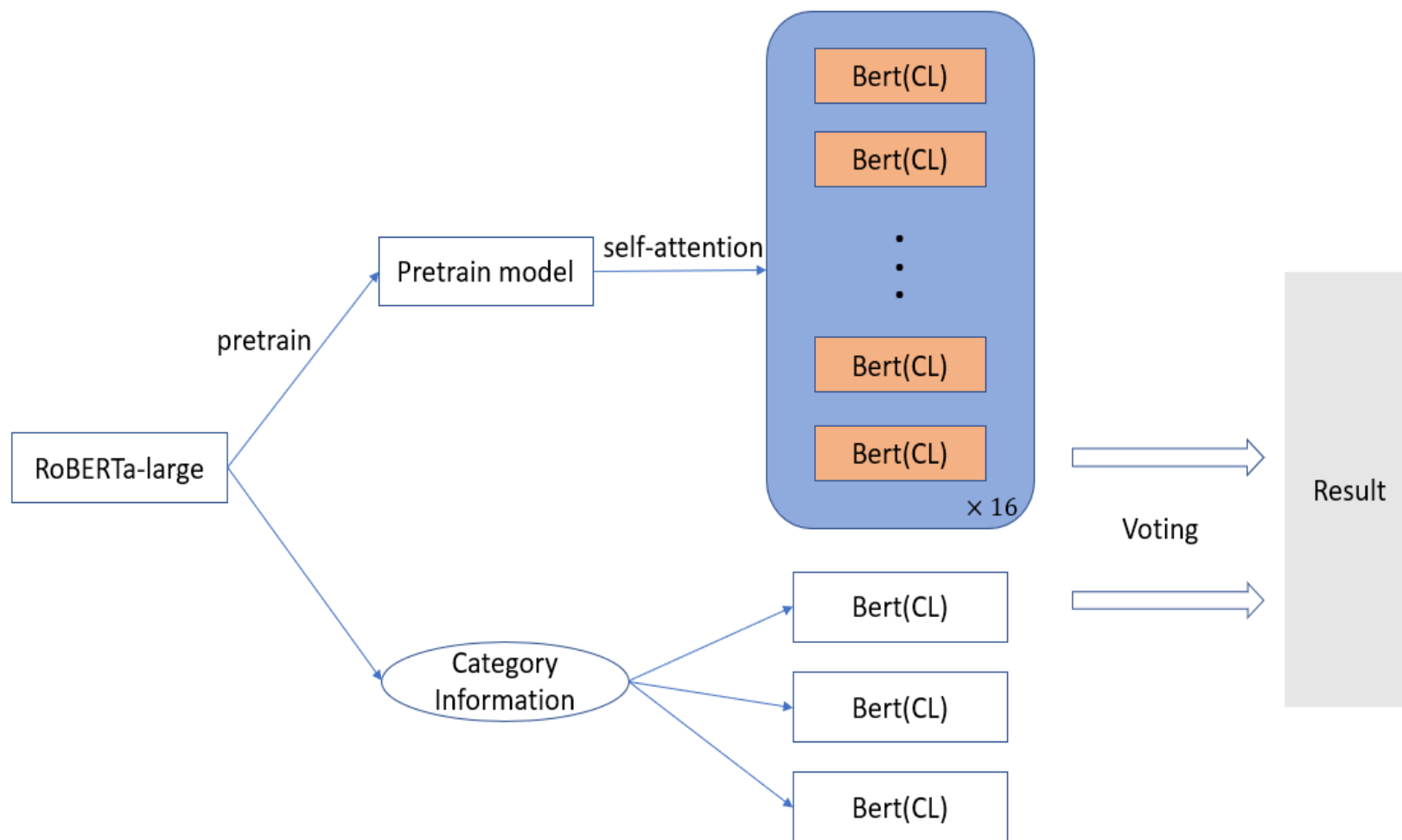
- P / R / F1

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$F1 = \frac{2\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

实验



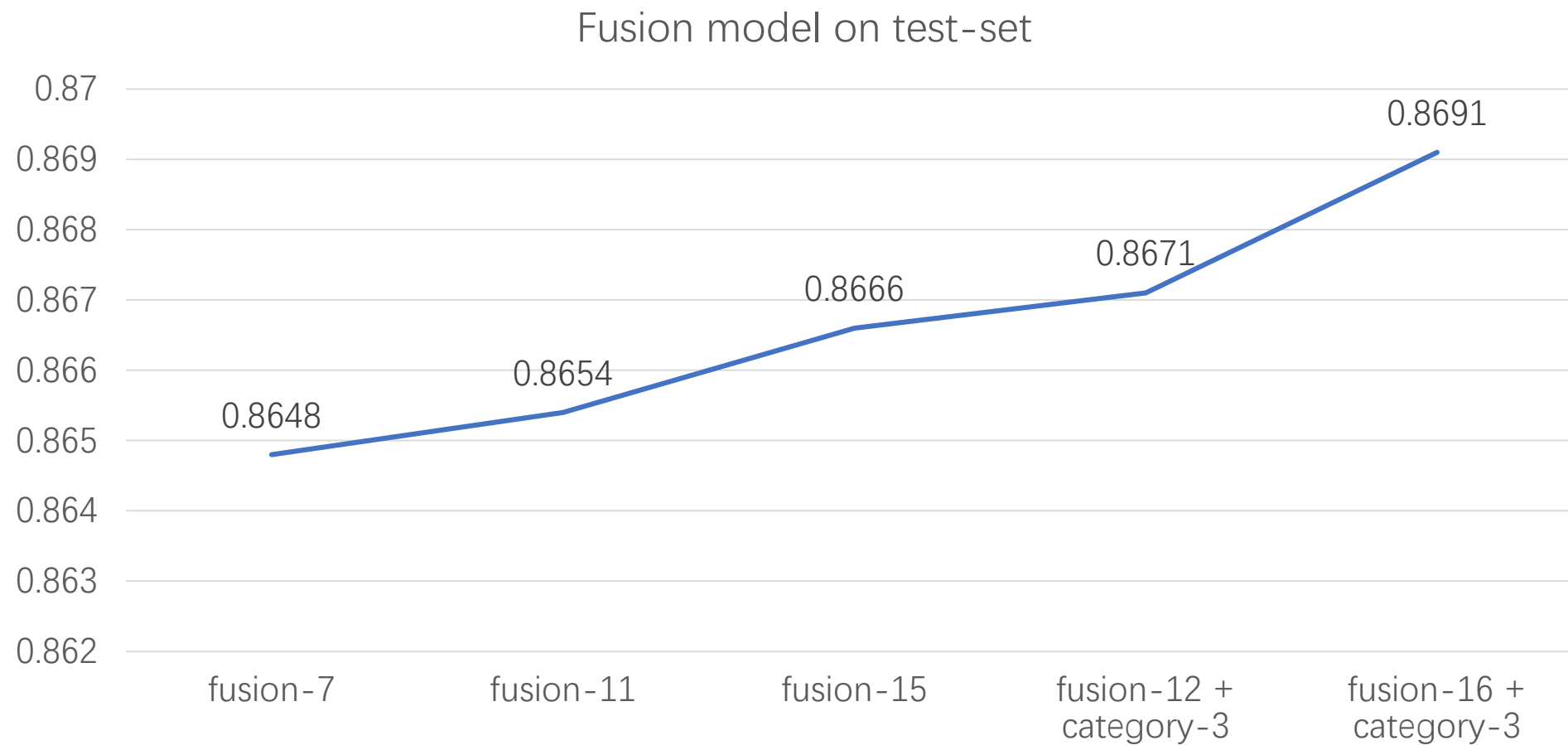
实验

Table 2. Performances of various single model on the validation set ↵

Model ↵	F1-score(%) ↵
ERNIE ↵	85.54 ↵
BERT ↵	85.63 ↵
BERT-wwm-ext ↵	86.51 ↵
BERT-wwm-ext(with category) ↵	86.47 ↵
RoBERTa-wwm-ext-large ↵	87.43 ↵
RoBERTa-wwm-ext-large(pre-trained) ↵	87.73 ↵

↵

实验



实验

- 单模型对比上，序列标注模型比阅读理解模型F1值高，但都能达到89+的结果，证明了预训练模型的有效性
- 使用并集和交集策略结合多个单模型的结果，能够提升性能
- 序列标注模型和阅读理解模型的结果可以互补，在线下验证集合得到最好的F1值

结论

- 我们使用预训练模型RoBERTa-Large WWM作为Baseline
- 通过对任务评测目标分析，我们采用了模型融合以及引入病种信息的方案，在最终的测试集上取得了第三名
- 简单地更改预训练模型的结构并不能带来明显的性能提升，说明如何利用预训练模型还是一个值得进一步探索的问题。

Thanks !