

中医辨证辨病及中药处方生成技术研究

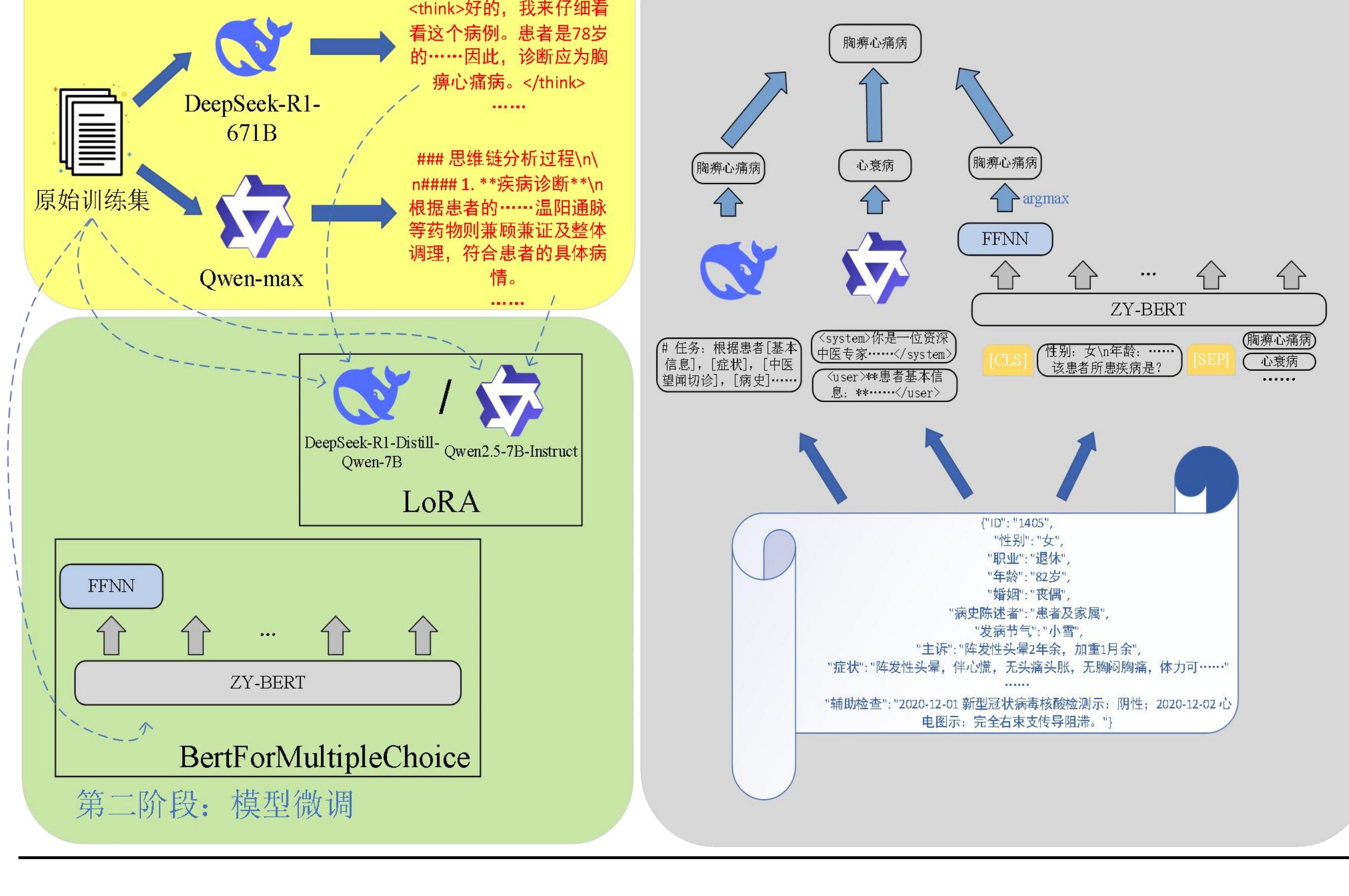
张剑昆 马龙龙 刘汇丹

张剑昆 zhangjiankun24@mails.ucas.ac.cn

研究背景

- 近年来，大语言模型在各个领域蓬勃发展，但在传统中医药领域的应用尚未获得充分关注。
- 辨证论治是中医认识和治疗疾病的核心原则和方法，通过综合、分析来自望、闻、切、诊的患者临床信息，归纳对应证型并辨明所患疾病，进而开具合适处方予以治疗。
- 基于CCL25-Eval评测任务展开实验，并最终获得该任务的三等奖。

技术方案



- 微调Qwen、R1和ZY-BERT，取三者预测输出的硬投票结果作为最终辨病结果。
- 收集来自Qwen-max模型的思维链数据以帮助模型显式学习相关知识。
- 收集671B版本R1的推理链以指导蒸馏模型的微调。
- 以四选项单项选择问题应用ZY-BERT模型。

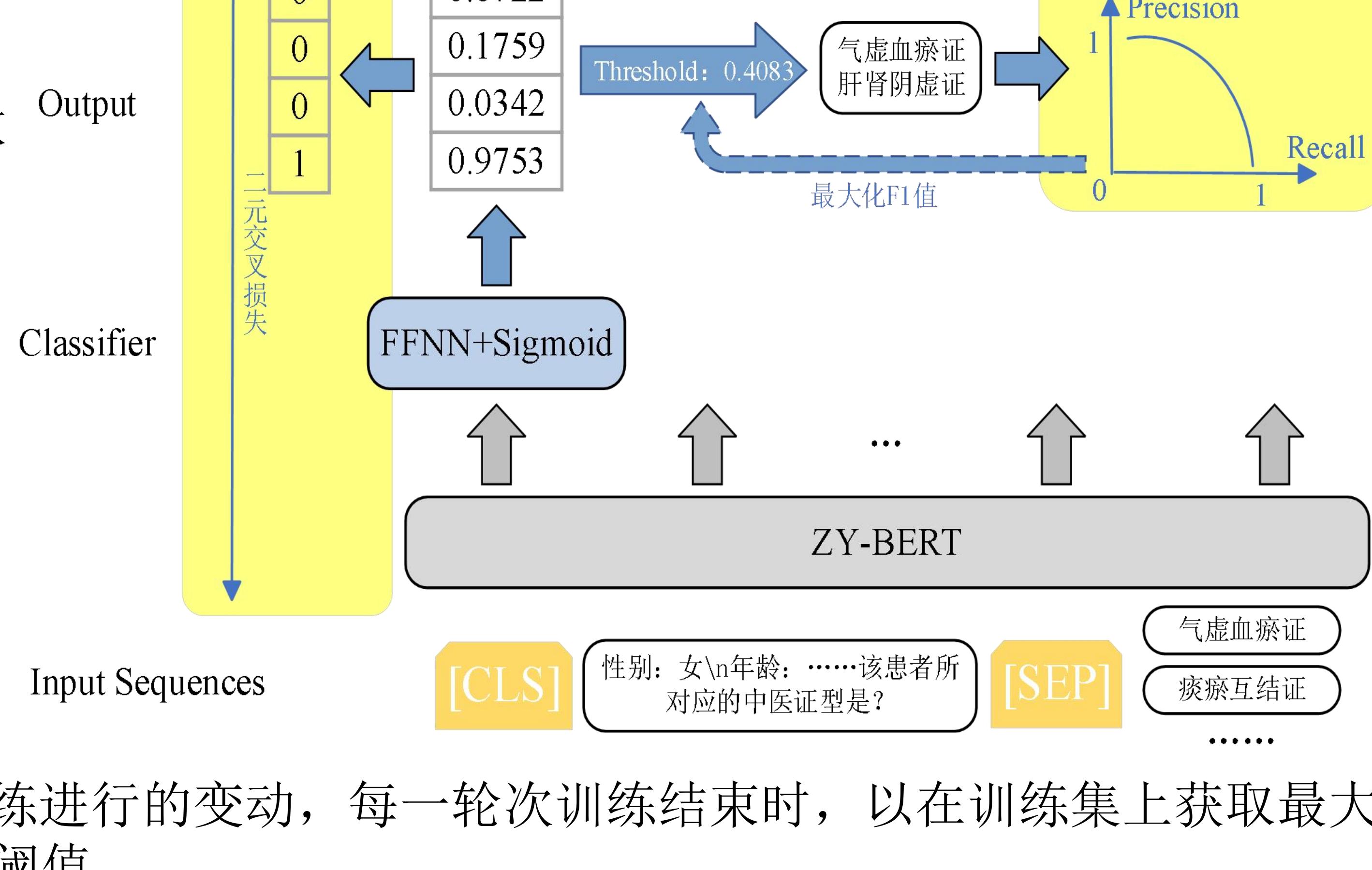
- 以ZY-BERT模型为基础，取[CLS]的最后一层输出传入前馈神经网络，后经Sigmoid激活函数处理输出患者症状表现属于该证型的概率，后由阈值判断机制获得最终辨证预测。
- 通过随机舍弃的方式将单个问题的总选项数控制在4，以避免错误的拟合（倾向于输出概率0）。

- 为适配模型能力随训练进行的变动，每一轮次训练结束时，以在训练集上获取最大F1值为标准调整概率阈值。

在处方生成任务上，通过对Qwen-7B-Instruct模型做提示词优化和有监督微调作为解决方案。

实验评估

模型	辨证辨病	处方生成
baseline	0.45	0.41
Qwen2.5-7B-Instruct	0.27	0.29
DeepSeek-R1-Distill-Qwen-7B	0.20	0.24
HuatuoGPT2	0.46	0.37
TCMLLM	-	0.33
TCM-Chat	-	0.38
ours	0.59	0.42



- 辨证与辨病结果准确率的平均值构成第一个分数；处方（视为药物列表）生成结果与标准答案的Jaccard相似度、F1分数和药物数量匹配程度的平均值构成第二个分数。
- 上述方法在模型原始表现的基础上取得极大提升同时，甚至超越了领域大模型在该数据集下的表现，也在辨证辨病任务上远超baseline结果。