

2020 Study REPORT



个人总结

蔡泽锋



Content

01 学期总结

02 金融文本的特点与任务

03 模型构思

04 主要步骤





/01

学期总结

论文与实验

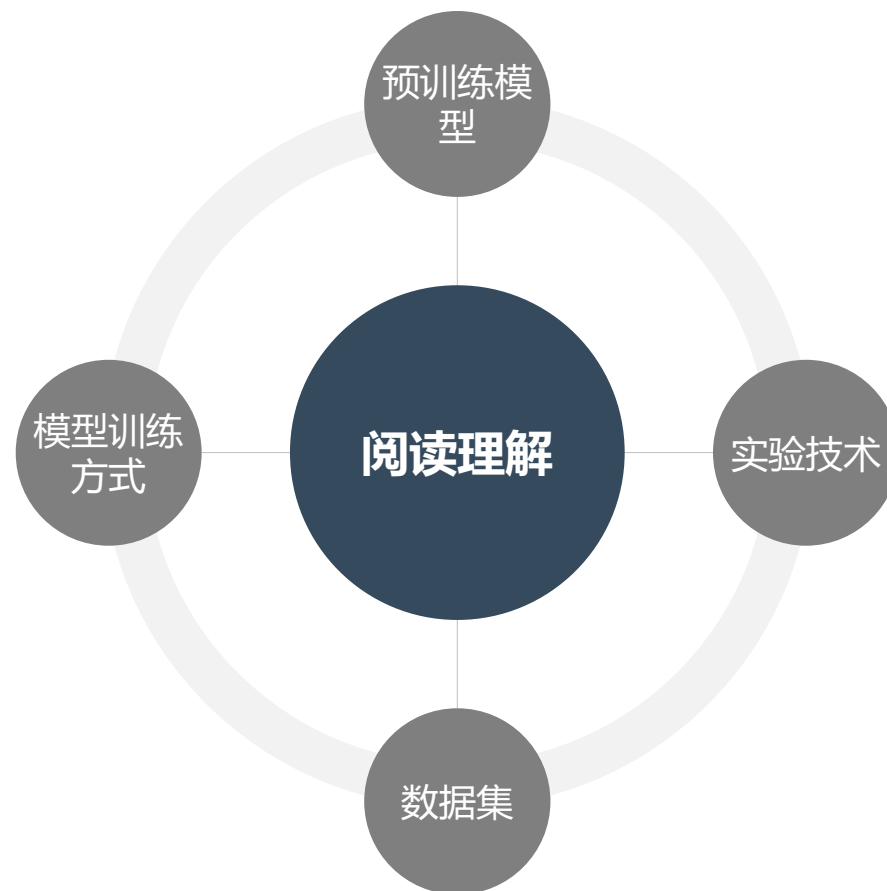
学期总结

预训练模型

阅读了5~10篇有关BERT和10+有关MRC的论文。

数据集

RACE, SQUAD, MRC数据集, 理解数据和实验结果。



训练方式

学习了知识蒸馏, 优化理论, 课程学习, 模型集成等知识。

实验技术

混合精度, 梯度截断, 梯度累加, checkpoint, AutoML。



/02

金融文本上的任务

情感分析, 对话问答



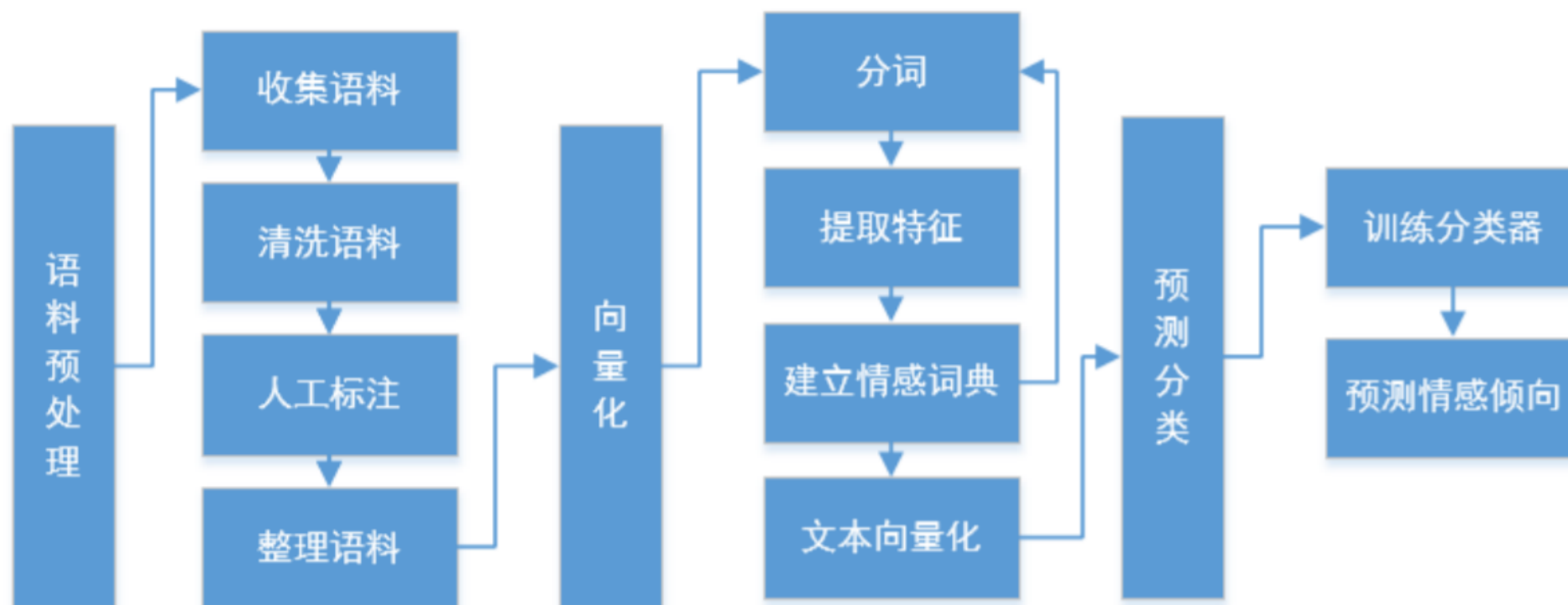
金融文本的特点

- 目前对互联网的金融文本评论研究还处于起步阶段。把针对商品评论和电影评论等传统的方法直接应用到金融领域，效果比较不理想。因为金融领域的文本与商品评论的特点不同，商品评论可以根据描述文本属性、特征的形容词、副词计算文本的特征，而金融文本的有许多的专业术语、词语的搭配、句式结构与日常生活接触的文本有一定的差异。通用的语料有通用的情感词典可用，但是金融评论文本并没有现成的情感词典。

金融文本上的情感分析

- 互联网上的金融信息的来源包括研报、公告、新闻、股吧等渠道，金融信息中会包含 投资者在投资中的情感和行为，D投资者的情感和行为可能会影响到股票的走向趋势，将基于金融新闻的情感量化后用于预测股价，准确率提高了9%。因此，网络金融信息与股票市场存在一定的相关性，可以通过研究这些信息的情感倾向进而提高预测股票市场形势的准确性。

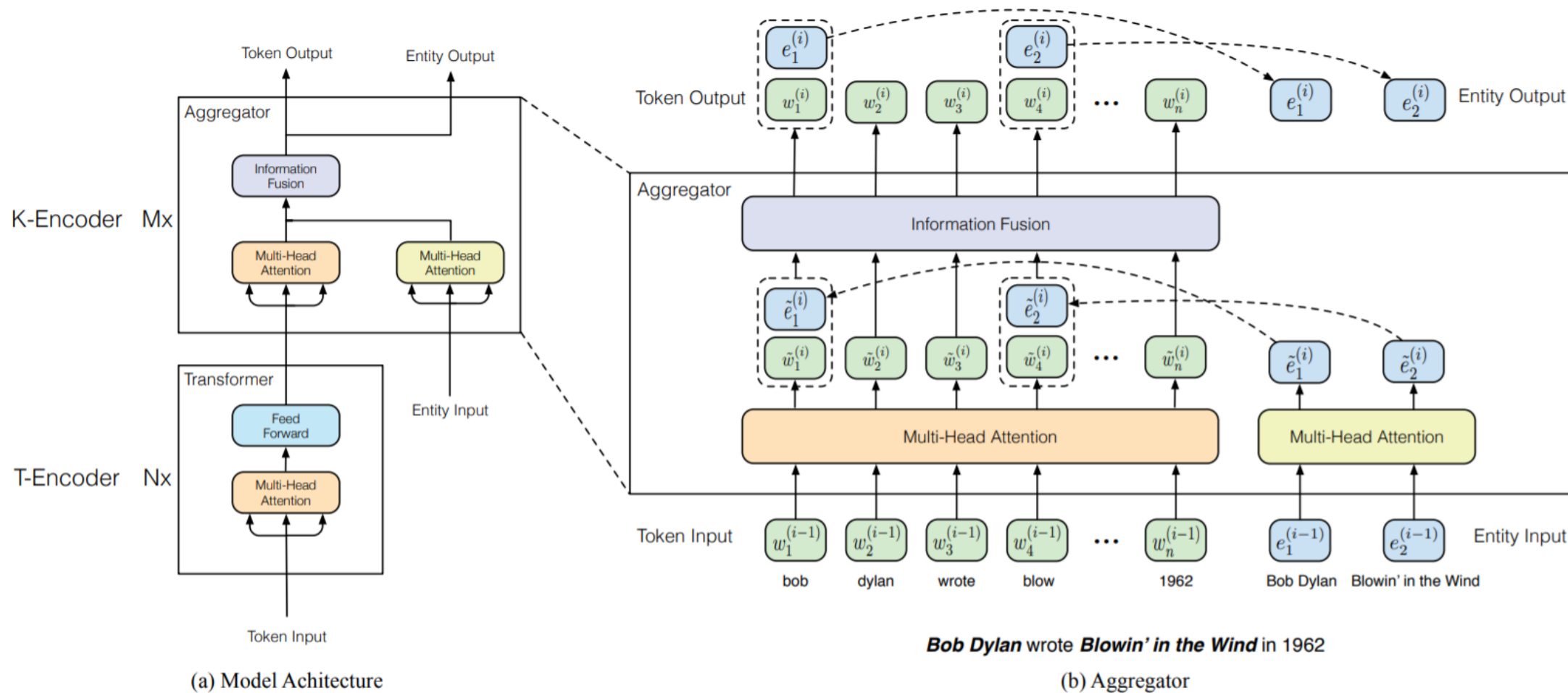
模型构思



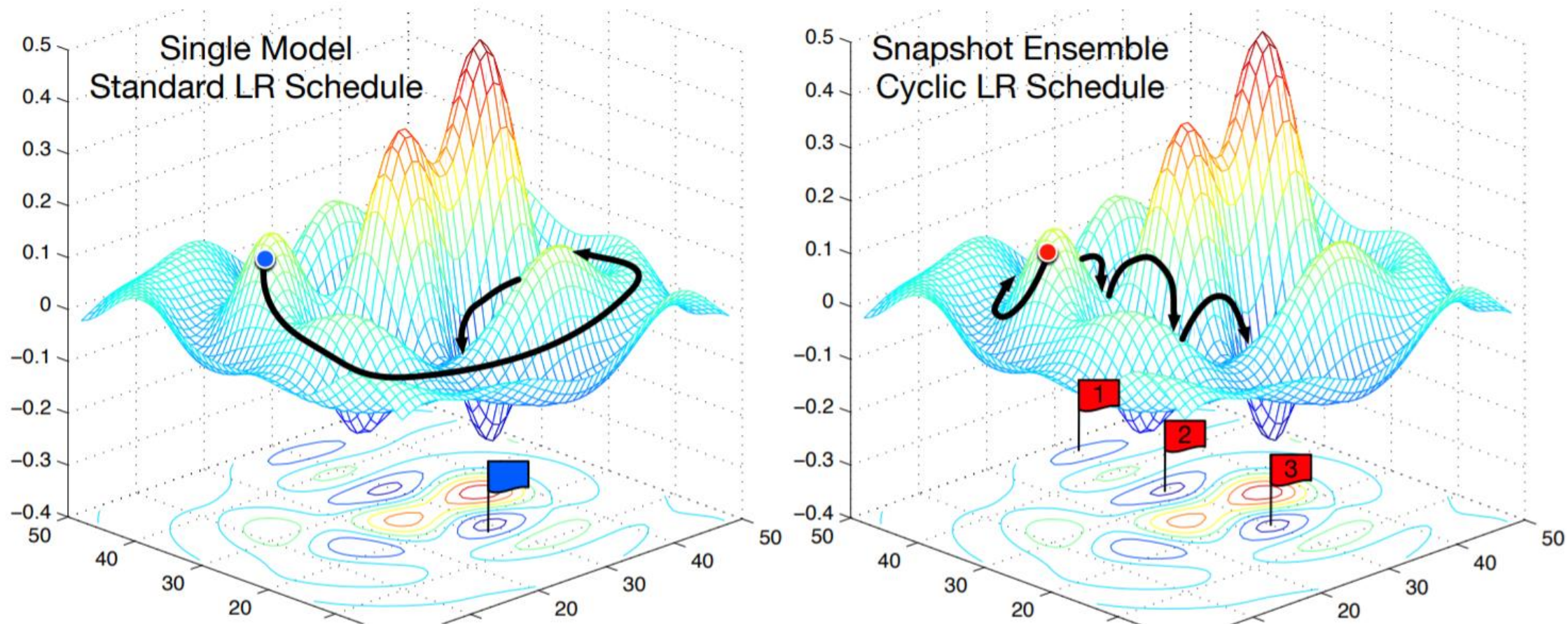
数据预处理

- 繁体转简体
 - 『精伍鴨頭湖北菜館』位於環球大廈一樓西側1054號，服務電: 88259300，位於鬧市區，不太好停車
 - 『精伍鸭头湖北菜馆』位於环球大厦一楼西侧1054号，服务电: 88259300，位於闹市区，不太好停车
 - 影响面 4.61% (4839/105000)
- 特殊表情字符不被切分
 - Emoji (pyemoji) 😊
 - 高频表情
点评 ^_^ 。

主要模型



模型集成





End

Thanks For Your Listening.

蔡泽锋-个人总结

2021.3.17



End

Thanks For Your Listening.

蔡泽锋-个人总结

2021.3.17



End

Thanks For Your Listening.

蔡泽锋-个人总结

2021.3.17



End

Thanks For Your Listening.

蔡泽锋-个人总结

2021.3.17



End

Thanks For Your Listening.

蔡泽锋-个人总结

2021.3.17



End

Thanks For Your Listening.

蔡泽锋-个人总结

2021.3.17