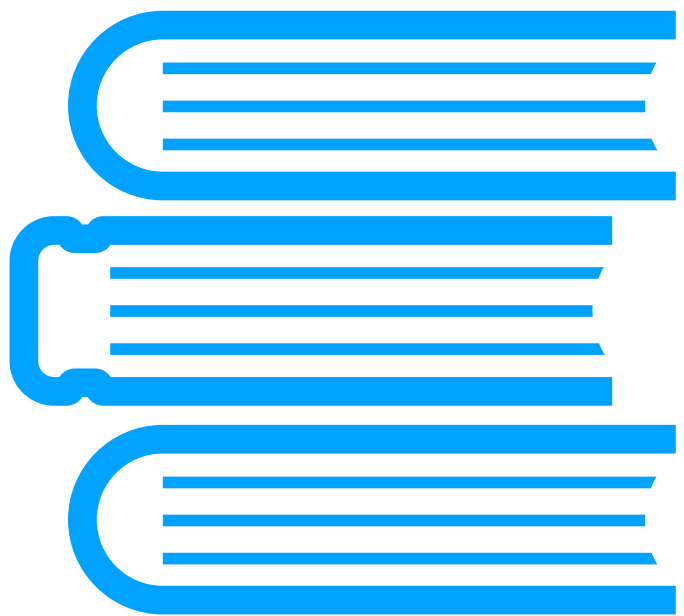


基于NLP的金融营销活动情感分析

队伍编号: 276

2017-10-21

目录



问题概述及分析理解

国内外研究总结

模型构建过程

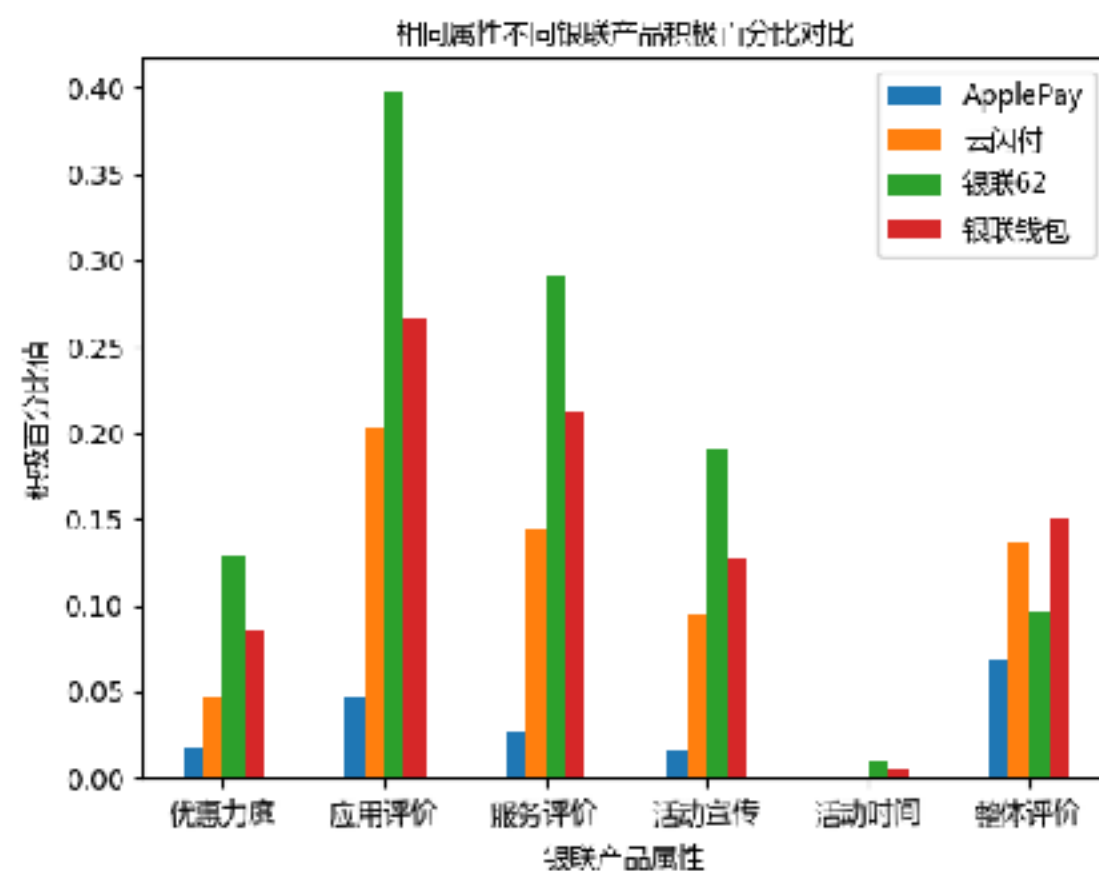
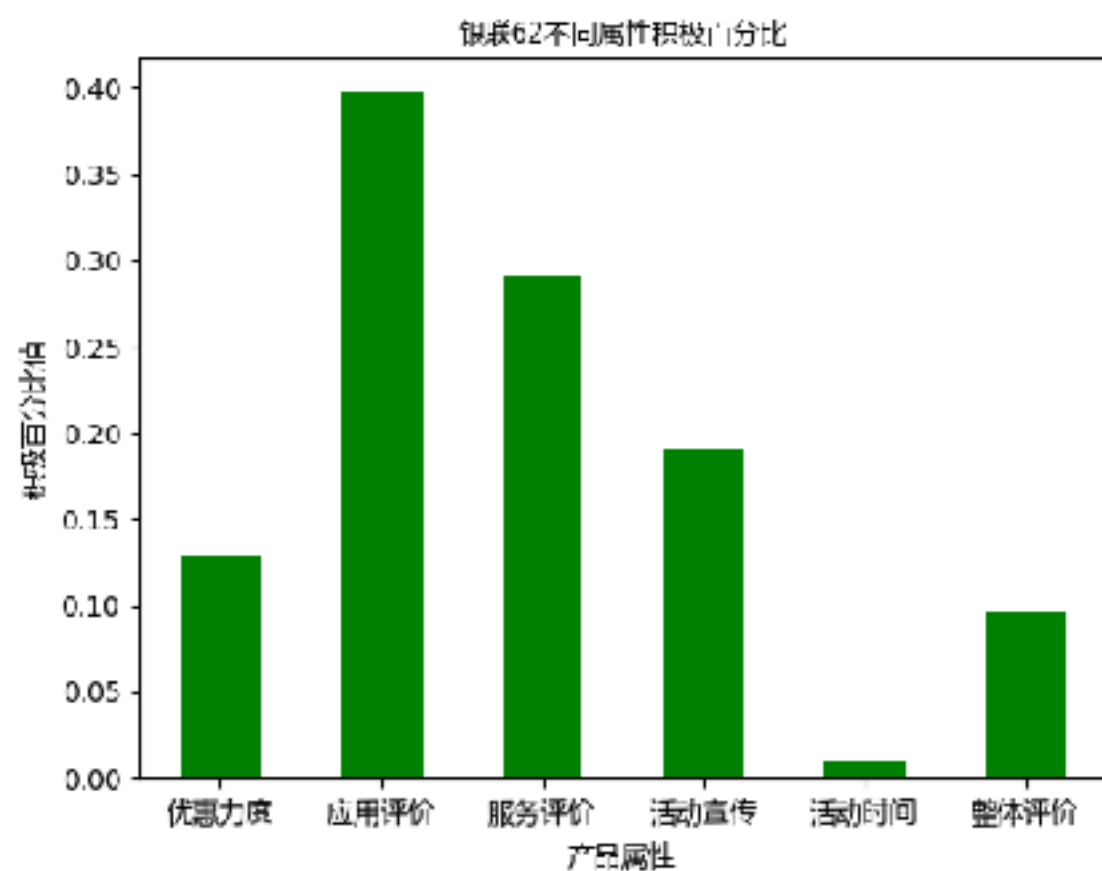
问题总结与优化建议

问题概述

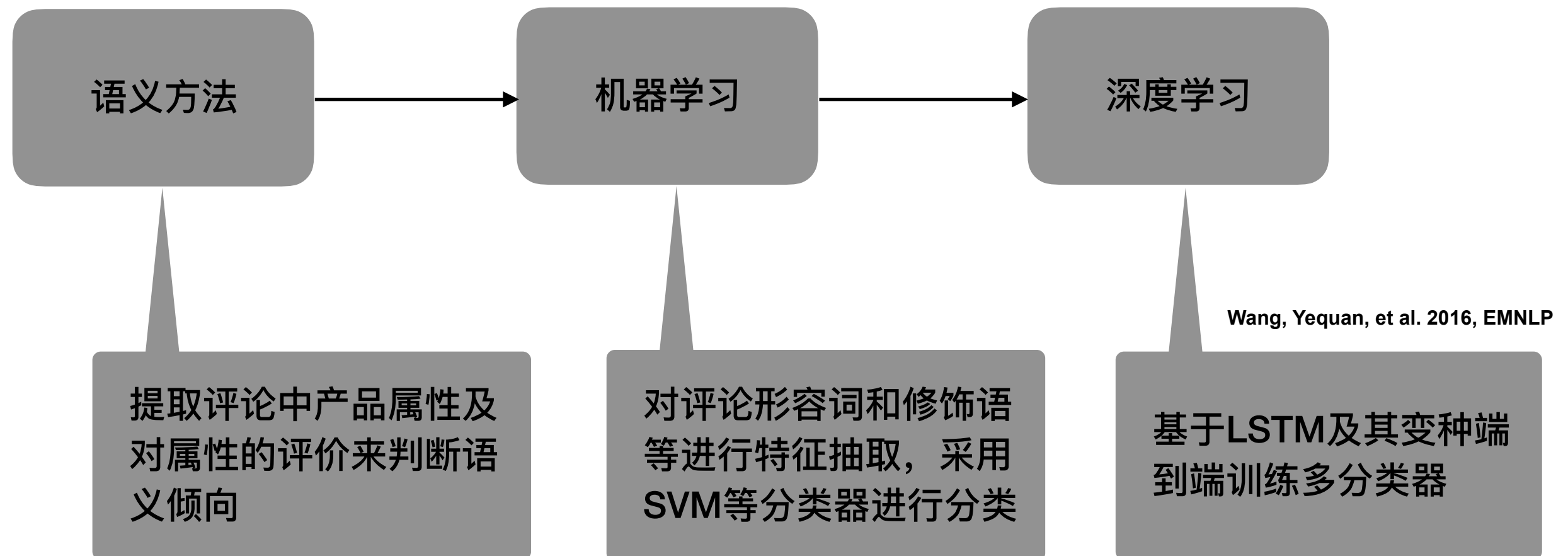
给定针对不同银联产品的评论文本及其对应的银联标签和商户标签，对不同银联产品推广效果进行评估和分析。

评论文本	活动内容	银联标签			商户标签		
评论内容	银联钱包	优惠力度	...	整体评价	服务质量	...	整体评价
	...						

积极百分比



国内外研究总结



数据选择

评论文本	活动内容	银联标签			商户标签		
评论内容	银联钱包	优惠力度	...	整体评价	服务质量	...	整体评价
	...						



评论文本	活动内容	银联标签-整体评价
评论内容	银联钱包	好/中/差
	...	

分析理解

- 模型：三分类
- 输入：评论文本，活动内容
- 输出：整体评价
- 评价指标：准确率，精确率，召回率和F1评分

数据预处理

☐ 原始数据的行列处理

- ☒ 第一批数据没有链接列，第二批数据评论文本标签列不齐。

☐ 异常值和空值处理

- ☒ 第一批训练数据中某活动内容列有3条值为银联标签。
- ☒ 测试集中出现银联62，银联62活动，ApplePay，Applepay，Apple_Pay。
- ☒ 训练样本83100的评论文本内容为空。

☐ 初赛和决赛训练数据对比

- ☒ 每个活动内容样本数量不变，但增加了每类活动内容的好和差的样本数量。

数据概貌(1/2)

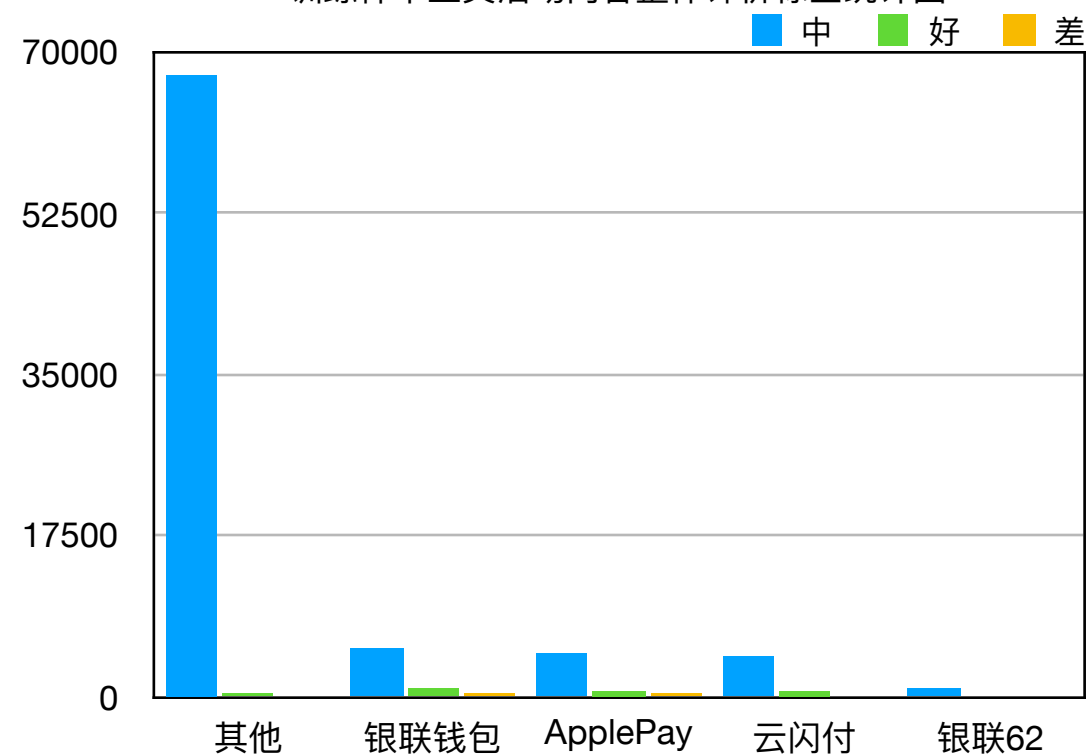
训练样本整体评价标签统计表

活动内容	中	好	差
其他	67528	337	35
银联钱包	5216	1004	434
ApplePay	4593	662	389
云闪付	4524	756	223
银联62	910	99	24

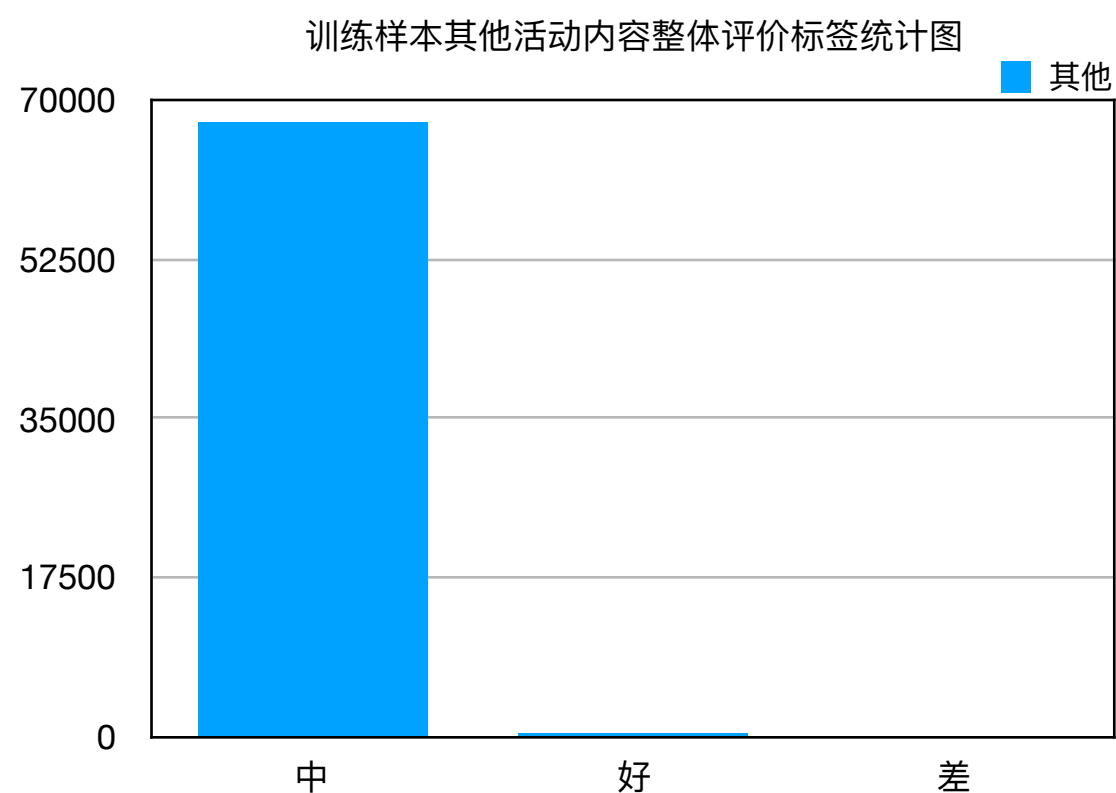
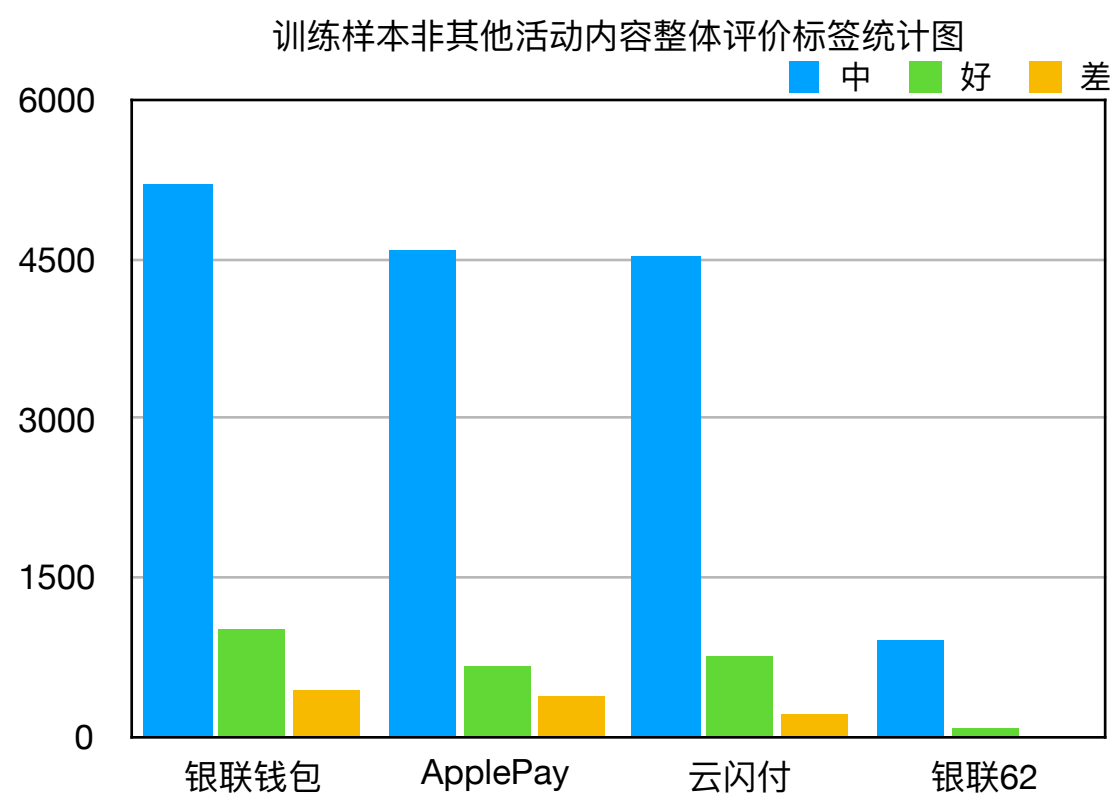
测试样本整体评价标签统计表

活动内容	中	好	差
其他	958	359	59
银联钱包	87	182	33
ApplePay	263	109	137
云闪付	627	288	106
银联62	53	47	16

训练样本五类活动内容整体评价标签统计图



数据概貌(2/2)



不平衡处理

☐ 采样方法

☒ 上采样: **overfitting**

☒ 下采样: 学习不充分

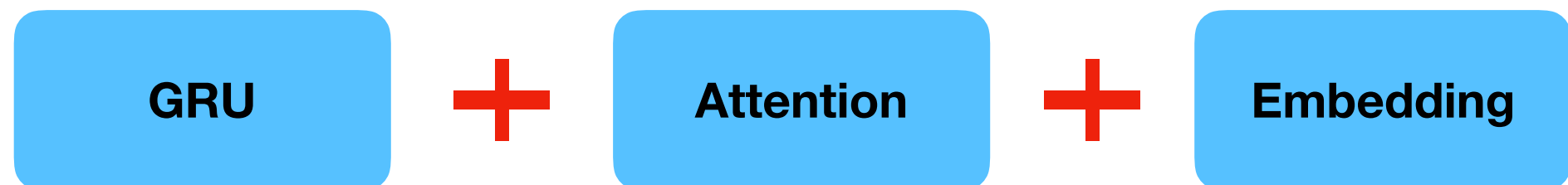
采样方案	好	中	差
方案1	固定值采样, 比例为1:1:1		
方案2	降采样中, 过采样差, 比例为1:1:1		
方案3	过采样差, 好, 比例为1:1:1		
方案4	参照中, 比例为1:2:1		

模型构建

☐ 数据集划分



☐ 模型构建



训练和预测

☐ 训练

- ☒ 针对每类活动内容，不平衡处理后，根据 bias/variance 均衡进行模型选择，最后对被选模型持久化保存。

☐ 预测

- ☒ 对每个待测样本，过滤评论文本内容，根据活动内容，选择相应模型进行预测。
-

实验结果

关键参数

batch_size	embedding_dim	drop_out	sequence_len	words_len	hidden_dim
32	64	0.3	25	1000	256

测试样本四个评价指标值

指标类型	中	好	差
精确率	0.81	0.9	0.93
召回率	0.96	0.67	0.6
F1评分	0.88	0.77	0.73
准确度	0.84		

问题总结与优化建议

- ☐ 其他不平衡方案尝试
- ☐ 利用银联标签下的非整体评价标签和商品标签
- ☐ 其他模型尝试
- ☐ 活动内容不分类下的训练和预测

TKS

请各位评委专家批评指正