

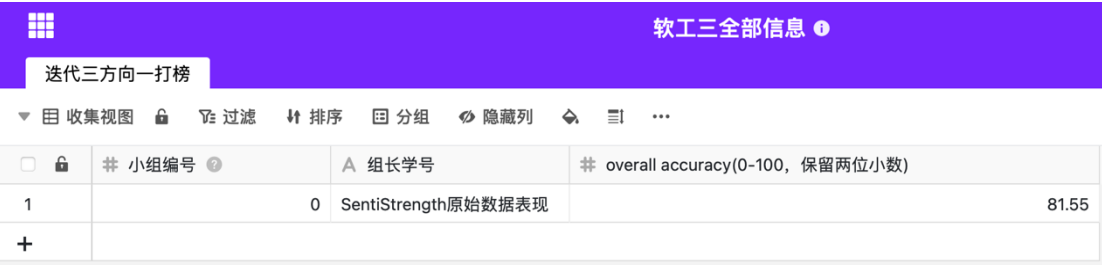
# 迭代三补充说明

2023.05.21

本文件将对软工三迭代三中方向一： 算法优化（规则式）以及方向二：算法优化（学习式）的打榜机制进行补充说明。


## 打榜方式

本次我们将通过调查问卷的形式进行打榜（方向一打榜链接<sup>1</sup>；方向二打榜链接<sup>2</sup>）。同学们需在问卷中填写组号，以及优化后的工具数据。我们将通过 overall accuracy、以及工具在三种类型的情绪极性上的精度（ $P$ ）、召回率（ $R$ ）、F-measure（ $F$ ）来衡量工具的表现。上述指标的计算方式可参考 SESSION 工具的论文<sup>[1]</sup>。其中，我们将以 overall accuracy 为主要的衡量指标，并根据它进行排名。**打榜问卷中仅需要填写 overall accuracy**（数据需精确到小数点后第四位，考虑到在线问卷会自动保留两位小数，请将原本的四位小数“0.xxxx”转化为“xx.xx”格式输入）。其他的细节数据需在后续的优化说明文档中给出。打榜过程中，同学们可通过问卷的统计结果了解其他同方向小组的优化表现。打榜结束后，我们将根据排名的 30%/40%/30%，进行+4 分 /+2 分 /+1 分的赋分。图一展示了填写小组数据时的大致界面。点击左上角的九宫格图表可进入主界面，通过主界面中的“方向一排名榜”和“方向二排名榜”可看到以更新的数据排名（如图表 2 所示。）



软工三全部信息 ⓘ			
迭代三方向一打榜			
▼ 收集视图	🔒	🔍 过滤	🔼 排序
📊 分组	👁️ 隐藏列	🔍	...
<input type="checkbox"/> 🔒	井 小组编号 ⓘ	A 组长学号	井 overall accuracy(0-100, 保留两位小数)
1	0	SentiStrength原始数据表现	81.55
+			

图表 1 填写小组数据的界面图



👤 2023软工三 ▼	👤 190
📊 全部小组 (共享)	📊 Gitlab 小组信息导出情况 (共享)
📊 迭代二中期答疑问题收集 (共享)	📊 方向一排行榜 (共享)
📊 方向二排行榜 (共享)	

图表 2 排名展示界面

<sup>1</sup> 数据收集表 方向一打榜, 点击链接填写 : <https://table.nju.edu.cn/dtable/collection-tables/41b3c0c3-b89e-4965-bae6-c51e03e35948>

<sup>2</sup> 数据收集表 方向二打榜, 点击链接填写 : <https://table.nju.edu.cn/dtable/collection-tables/746f1fe2-89b4-4d0b-9446-9bff36b536a6>

## • 评估数据集

对于方向一：算法优化（规则式），我们将在完整的 SOF4423 数据集上评估工具的表现（该数据集已在迭代二中发布）。**同学们需在问卷中填写算法在完整的 SOF4423 数据集上的表现。**选择方向一的小组需公布优化后的代码和词典，以及最终数据。**对于方向二：算法优化（学习式），我们将发布已经划分好的 SOF4423 数据集（本次会新发布划分好的数据集）。**选择方向二的小组仅可以在训练集上训练算法，**在问卷中填写算法在测试集上的表现即可。**选择方向二的小组需公布训练参数、训练日志、最终模型、最终数据。助教保留重新训练和验证模型的权利。**在排行榜上伪造结果，以及将测试集用于训练的行为都将被视为作弊。对于存在作弊行为的小组，我们将直接扣除 41% 的总评分数。**

选择方向一、方向二的小组需撰写优化说明文档，**文档中需明确给出小组最终的细节评估结果**，格式如图 3 所示。**对于方向一**，需说明新增了哪些规则，并为新增规则举出合适的例子，解释其是如何生效的；若优化了词典，需要说明采取了哪些操作，是如何进行优化的；并进行合适的消融实验，去分析新增的规则或者词典优化的效果。具体分析同样可参考 SESSION<sup>[1]</sup> 以及 SentiStrength 论文<sup>[2]</sup>。**对于方向二**，需说明说明算法训练的整体流程，如采取了哪些数据处理办法、或数据增强操作、或尝试过哪些算法等。若尝试过多种分类算法或数据处理操作，可在优化说明文档进行说明并展示实验数据。方向二同样可进行合适的消融实验，去分析整体算法中各部分的效果，具体分析可参考 SentiCR<sup>[3]</sup> 和 Senti4SD 论文<sup>[4]</sup>。

Tool	overall accuracy	positive			neutral			negative		
		P	R	F	P	R	F	P	R	F
SentiStrength	81.55%	88.90%	92.34%	0.906	92.76%	63.58%	0.754	66.83%	93.18%	0.778

图表 3 细节评估结果展示图（该数据摘自 SESSSION 论文）

## • Reference

- [1] Sun K, Gao H, Kuang H, et al. Exploiting the unique expression for improved sentiment analysis in software engineering text[C]//2021 IEEE/ACM 29th International Conference on Program Comprehension (ICPC). IEEE, 2021: 149-159.
- [2] THELWALL M, BUCKLEY K, PALTOGLOU G, et al. Sentiment strength detection in short informal text[J]. Journal of the American society for information science and technology, 2010, 61(12): 2544 – 2558.
- [3] Ahmed T, Bosu A, Iqbal A, et al. SentiCR: a customized sentiment analysis tool for code review interactions[C]//2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, 2017: 106-111.
- [4] Calefato F, Lanubile F, Maiorano F, et al. Sentiment polarity detection for software development[C]//Proceedings of the 40th International Conference on Software Engineering. 2018: 128-128.