

FewCLUE: 中文小样本学习—调查问卷反馈信息

CLUE Benchmark, CLUEbenchmark@163.com, 2021-05

一、调查背景

本次调查问卷主要针对任务和数据集设定进行调查，调查对象为FewCLUE小样本学习交流群，获得35份有效样本。

二、调查问卷主要反馈

- 1) 希望任务丰富多样、适应真实场景；
- 2) 数据集数量: 提供9个左右数据集
- 3) 单个任务数量: 按类别采样16为主；
- 4) 半监督学习: 小样本测试还应提供大量无标签数据
- 5) 测试零样本学习

三、调查问卷主要信息汇总

数据集数量、样本数量、采样方式、无标签样本、零样本学习、任务类型和特点

反馈一致的问题

1、问题（数据集数量）：一个具体的数据集或任务，应该包括多少数据量？

答案汇总：提供8-10个数据集。

具体信息：

- 1) 8-10个之间，占比69%；
- 2) 5-6个占比23%。

2、问题（样本量）：按照类别进行采样获得了多少答案？

答案汇总：按类别进行采样为主，如16个。

具体信息：

- 1) 每个类别采样16个占63%（数据未归一化）；
- 2) 每个类别采样8个占26%；
- 3) 训练集总数量32占14%；

注：需要注意的是对于2个类别，采样16个和总样本数量32个效果是一致的。

3、问题（样本量）：对于类别比较多的任务，如类别数为10、50个，应该包括多少总的样本数量？

答案汇总：按类别采样。

采样16个（40%）；

采样8个（26%）

4、问题（无标签样本）：对于一个特定任务，除少量标注数据外，是否应该提供大量无标注数据？

答案汇总：应提供大量无标签数据。

具体信息：

应该提供，因为真实场景也有大量无标签数据（占77%）；

不需要提供，目标只是少样本学习（23%）

5、问题（零样本学习）：零样本学习，以GPT3为代表，近年来有很多创新应用和发展。

FewCLUE是否应测试零样本学习？

答案汇总：应测试零样本学习。

具体信息：

可以。零样本学习效果可以作为对比，以及可以考察生成类模型的能力（占69%）；

不需要。确保目标只是小样本学习（31%）

6、问题（任务类型和特点）：你认为好的中文小样本学习数据集应该是怎么样的？

请告诉我们你的想法，最好提供依据（主观题）。

答案汇总：任务丰富多样、适应真实场景，占比（63%）

7、问题（建议）：对中文小样本学习测评的建议和意见？

答案汇总：

希望明确基础预训练模型；

保证小样本对等的公平性；

不鼓励太多trick参赛；

综合考虑方案难度和复杂化等。

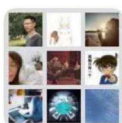
反馈不明确的问题

1、问题：对于一个具体的任务，应该提供多少份训练和验证集？

答案汇总：提供多份 vs 提供1份

每个任务5份训练集，以及对应的验证集，确保结果稳定可靠（占52%）；

每个任务一份训练集，以及对应的一份验证集，简单方便（占48%）；



FewCLUE: 中文小样本学习交流群

