

1. 测试目的

本阶段测试的主要目的是：

1. 测试对于有解决方案的数据，对于将不同比例的数据以不同的方式放入训练集、知识库、测试集后得到的模型准确性。
2. 测试上述方法中的模型处理时间是否能满足要求。

2. 只有 500 条有标签数据的情况

2.1 对于微调数据，不将其放入知识库的方法

- 方案 1：按照训练数据：测试数据 = 325 : 175 的比例测试。
方案 2：按照训练数据：测试数据 = 300 : 200 的比例测试。
方案 3：按照训练数据：测试数据 = 350 : 150 的比例测试。
方案 4：按照训练数据：测试数据 = 375 : 125 的比例测试。
方案 5：按照训练数据：测试数据 = 400 : 100 的比例测试。

2.2 对于微调数据，将其全部放入知识库的方法

同 2.1 的五种方案中的训练数据与测试数据的比例，进行测试，与 2.1 的区别仅在于我们将微调使用到的数据全部放入到了外挂的知识库里面。

2.3 对于微调数据，将其部分放入知识库的方法

五种比例仍同 2.1。对于五种比例的方案，分别再测试将微调数据的 25%、50%、75% 放入知识库中的模型精度。

3. 2000 条数据标签之后的情况

3.1 数据情况

2000 条数据补全后，共有 2000 条有监督数据集 A。500 条有监督数据集 B。
数据分配给三部分：微调训练集、微调测试集、外挂知识库。

3.2 测试方法

方案 1:

其中数据集 A 按照 70%训练集, 30%测试集分配 (可浮动)。数据集 A 中用来作为训练集的 70%和数据集 B, 共同作为外挂知识库。

方案 2:

数据集 A、B 混合, 按照 70%训练集, 30%测试集分配 (可浮动)。其中用来作为训练集的 70%有监督数据作为外挂知识库。

方案 3:

数据集 A 与数据集 B 中的 300 条 (可调整) 混合。按照 70%训练集, 30%测试集分配 (可浮动)。数据集 B 中剩余的 200 条与混合后的 70%训练集, 共同作为外挂知识库。