

1 Modify MLM Task

hard label is human annotated, a labor-intensive and time-consuming task. When training, we use cross entropy loss to compare the hard label and the predicted soft label. Is there a gap? I have tried to use `torch.nn.CrossEntropyLoss()` to calc the ground truth and the predicted logit.

the problem is how to get the appropriate soft label?

2