

模型说明：

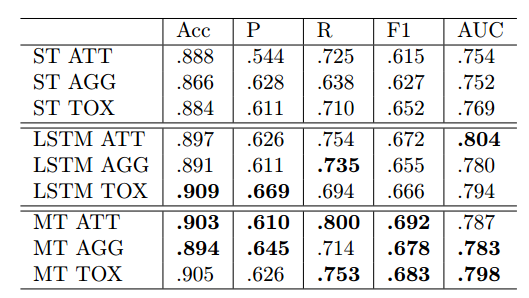
1. BERT作为encoder，得到输入单词的特征表示和句子级[cls]表示
2. (1) 将Aggression,Attack,Toxicity三个类别分别做label embedding

(2) 对于每个task,将label embedding和单词特征表示做attention,得到

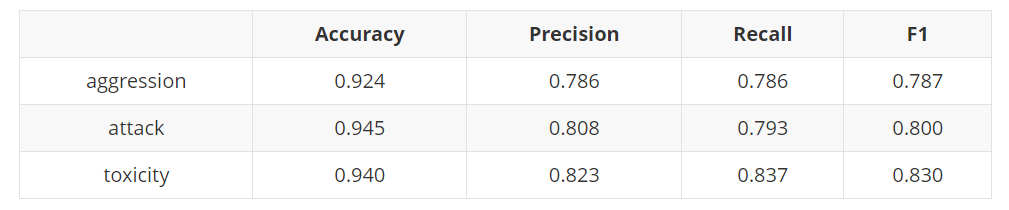
其中，为encoder输出，为label embedding

1. 对于每个任务，将不同任务得到的分别与[cls]表示拼接，之后softmax进行分类
2. Loss=0.3\*task1\_loss+0.3\*task2\_loss+0.3\*task3\_loss

之前论文中的结果



Bert Single task score



Bert Multi task score

