大家好。本组成员有，xxxxxx。我是xxx。由我汇报本组成果。

首先是数据预处理的部分。本次作业中，数据集的格式为：问题 + 句子 + 0/1标签，一个问题对应一篇文章，而这篇文章被拆分为许多分句，句子中是否含有答案用01标签标注。尽管是问答问题，但数据集的形式具有分类任务的形式。经过考虑，我们优先考虑用基于问答的方式来处理数据，也就是在用于SQuAD的BERT上fine-tune模型。

SQuAD 数据集的格式为文章，问题，答案三元组。为了将数据整合成这种形式，我们做了如下预处理，也就是将每个分句拼成文段，并构造三元组。下面是一个例子。

我们选择问答任务的原因为：

1. 作业数据集是将同一篇文章拆分成许多句子，这些句子之间的顺序可能会给答案作出贡献；

2. 问题文本提供了对答案范围的限制，如人名、地名、时间这些答案类型；但在缺乏全文信息的情况下，可能有“张冠李戴”的情况，选出并不符合问题信息的同类答案。

为了对比考虑全文的预训练模型与拆分成单句的预训练模型效果，我们也用分类任务的BERT模型进行了fine-tuning。

下面为模型结构的介绍：

后处理部分：

分类任务是容易的，对于问答任务，我们发现一些问题的答案在文章中出现多次。SQuAD模型原本从文章中输出一段的输出方式，不适合答案多次出现的情况。因此，在预训练模型输出答案片段后，我们与所有句子做相似度比对，这样就能得到多个答案的情况。

结果：在验证集上，问答模型略低于分类模型。对结果进行比对，我们发现分类模型的优势是能够发现文章中多处答案。缺点是由于数据不平均，对于某些问题将每个句子都分类成0。问答任务的缺陷是，只输出一个答案段，后处理时由于答案表述的不同，可能无法通过相似度、找子串的方法找出全部答案。

另外我们还尝试了其他模型，如XLNet，将数据做如下预处理。模型实现的方法是，使用开源的中文XLNet预训练模型，外加两个Dense层，一个bn层，训练后输出分类结果。