1.各位评委老师好，本次比赛中我们组的任务是对基于深度神经网络的模型应用——完形填空挑战模型进行优化。

2.接下来，我从以下几个方面介绍进行一一介绍。

3.近年来，tramsformer，bert模型的出现让NLP领域有了较大的发展，通过设计深度神经网络，实现智能理解人类语言一直是一种备受关注的方法。

4.本次优化的模型主要用于预测英语考试中完型填空的答案。该模型是基于深度神经网络构建而成，采取多种策略优化，并使用软投票方式整合多种模型，最终在语言考试任务的完形填空测试中可以得到高精度的结果，较好的完成了任务。

5.硬件环境中，我们使用了十核心的Intel CPU和NVDIA RTX 3090 24G版GPU，软件环境中，我们基于Linux系统，PyTorch框架，利用了transformer、pillow和deepspeed等包搭建了我们的应用模型。

6.模型的结构和流程如下，模型由编码器和解码器组成，编码器部分是由三种预训练模型，解码器由一个全连接层、一个gelu激活函数层、一个归一化层和一个全连接层组成。

7.文件结构如下图所示

8.模型训练中用到的优化方法包括算法优化和并行优化，其中，算法优化主要是梯度累加、数据增强、对抗训练和模型融合。并行优化主要使用数据并行、分布式数据并行方法和deepspeed框架。

9.批量的选取影响着模型的训练效果，但由于显存不足，在训练过程中我们一次只能读取少量的样本，导致模型收敛速度慢。因此我们采用梯度累加策略，通过先累加多个批量的梯度再更新对应参数的优化方法来获得更好的批量处理大小效果。

11.针对模型容易出现过拟合的问题，我们采用EDA方法进行数据增强，通过以下四种方法来增加数据噪声。如下图所示，使用EDA的测试精度比优化前的测试精度要高。

12.对抗训练是增强神经网络鲁棒性的重要方式，为了提高模型的泛化性能，我们使用了Goodfellow在2017年提出的快速梯度法来规范参数。由下图我们可以看出，同时使用EDA和FGM要比单独使用EDA时的测试精度要高

13.通过查阅相关文献资料，我们发现将多个弱学习能力模型集成到一个强学习能力的模型上可以得到更好的性能表现和泛化能力。在这个任务中，我们使用了三个预训练模型进行五折交叉验证训练，最终的预测结果使用软投票的方式进行模型融合。

14.由于数量过大，使用单卡训练一个epoch需要较长时间，于是我们使用Data Parallel方法，将一个batchsize的输入数据均分到多个GPU上分别计算，该方法实现简单，代码迁移成本低。

15.另一种优化方法——分布式数据并行DDP则依靠多进程来实现数据并行，这种方法既支持单机多卡也支持多机多卡。且相对DP，它通过Ruduce-Ring降低了I/O带宽。

16.而当模型大于显卡显存大小时，DDP无法使用，因此，我们使用DeepSpeed将模型参数拆分分散到各个GPU上，以实现大规模的计算，弥补了DDP的缺点。即，可使用更少的GPU训练更大的模型，而且不受限于显存。在本次任务中，我们即使用了原始的DeepSpeed训练框架，只进行简单的分布式训练，又启用了DeepSpeed训练框架下的Zero3策略，在保证精度的前提下，进一步提高了训练速度。

17.对于每个预训练模型，我们使用左图中所示的超参数进行训练。其中，单个模型Albert-xxlarge-v2在验证集上有最好的平均准确率，达到了90.42%。且模型融合后的最终整合结果为90.97%。

18.该模型不仅可应用在英文文本的完形填空，只要拥有适当的训练集，也可应用于其他语言，如中文等。除此之外，该模型还可以应用于文本纠错，通过将文本中错误的单词标记MASK，并进行预测MASK。这可广泛应用于文档或书籍的发表出版的审核，并可实现自动纠错。

18.我的演讲结束，感谢各位评委观看。