

原文作者：TanyaMalhotra

来源：Marktechpost

编译：DeFi之道

图片来源：由 Maze AI 生成

近年来，大型语言模型(LLM)在世界范围内受到了广泛的重视，在自然语言处理领域非常流行。。这使得我们能够以比以往更好、更清晰的`语言理解`来描述智能系统。

GPT-3、T5、PaLM等LLM的性能。已经有了明显的改善。而这些模型将继续存在，因为它们可以做任何事情，从通过学习阅读来模仿人类，到生成文本和总结大段文字。根据一些深入的研究，如果LLM很大，它的性能会很好。。通过在大量数据上训练这些模型，它们可以理解人类语言的语法、语义和语用。

openai开发的流行大型语言模型ChatGPT之所以发展如此之快。正是因为采用了人类反馈强化学习(RLHF)等先进技术。通过RLHF，机器学习算法结合人工输入提高了模型的性能。它针对预先训练的LLM进行了微调，LLM用于开发聊天机器人、虚拟助理和其他任务。

此外，基于ChatGPT等LLMs的预训练基础模型也有了明显的改进。这主要是因为三个变化：

1. 实践证明，模型的缩放对提高其性能非常有帮助。。以Pathways语言模型(PaLM)为例，该模型通过扩展少击学习极大地影响了其性能。小样本学习可以减少根据特定应用调整模型所需的特定任务训练样本的数量。通过使用Pathways语言模型，在6144TPUv4芯片上扩展和训练了5400亿个参数。PaLM显示了重复扩展的好处，超越了各种传统模式，显示了很大的进步。因此，深度和宽度的扩展是提高基本模型性能的重要因素。

2. 另一个变化是在预训练期间增加标记数量的过程。。像Chinchilla(开源语言模型)这样的模型已经证明，通过添加预训练数据，大型语言模型的性能会更好。

龙猫是一个计算优化模型。在同样的计算预算下在对70B参数和比Gopher模型多4倍的数据进行训练后，Chinchilla的表现一直优于Gopher南甚至比《GPT3》、《侏罗纪1》和《威震天图灵NLG》等LLM还要好。这清楚地描述了对

于每个计算上最优的训练，标记的数量应该相应地缩放——，即模型大小的两倍。所以训练标志的数量要翻倍。

3. 第三个变化是使用干净多样的预训练数据。卡拉狄加的表现证明了这一点。它是存储、混合和推理科学知识的大规模语言模型。。经过几篇科学论文的训练，卡拉狄加的性能比GPT3号和龙猫好。另一个大规模语言模型BioMedLM是用于生物医学文本的特定领域LLM。当训练特定领域的的数据时，它显示出很大的性能改进。它清楚地表明，对特定领域的数据进行预训练比对一般数据进行训练更好。LLMs的成功无疑归因于许多因素的混合。，包括使用RLHF和开发训练前的基本模型。这三个变化极大地影响了LLMs的性能。此外GLaM(CommonLanguageModel)通过使用稀疏激活的混合专家架构，以较少的训练成本扩展了模型的容量，从而显著提高了性能。因此这些变化为更高级的语言模型开辟了道路，这将继续使我们的生活更容易。