

# Practical Training on the Application of Pangu Big Mode

Haiying Yan Siwei Yan Xiuyong Hu

Shenzhen Aiwei Artificial Intelligence Co., Ltd., Shenzhen, Guangdong, 518000, China

## Abstract

With the rapid development of large-scale language models, Pangu big mode has demonstrated excellent application potential in multiple fields thanks to its powerful natural language processing capabilities. Starting from the architecture and technical characteristics, this paper explores the application of the Pangu big mode in different scenarios. Through systematic training design, a training environment was constructed, specific tasks were formulated, covering key aspects such as model pre training, fine-tuning, and compression acceleration. The performance of the model in practical applications was analyzed and summarized, aiming to provide technical references for the practical application of large-scale models.

## Keywords

Pangu big mode; pre training; fine tuning; model compression; practical training

## 盘古大模型应用实训

颜海鹰 颜思威 胡修勇

深圳爱微人工智能有限公司, 中国·广东 深圳 518000

## 摘要

随着大规模语言模型的迅猛发展, 凭借强大的自然语言处理能力, 盘古大模型在多个领域展示了卓越的应用潜力。论文从架构和技术特点出发, 对盘古大模型在不同场景下的应用进行探索。通过系统化的实训设计, 构建了实训环境, 制定了具体任务, 涵盖了模型的预训练、微调、压缩加速等关键环节, 并对模型在实际应用中的表现进行分析和总结, 旨在为大模型应用实践提供技术参考。

## 关键词

盘古大模型; 预训练; 微调; 模型压缩; 实训

## 1 引言

随着深度学习技术的飞速发展, 在自然语言处理(NLP)领域, 大规模的预训练模型取得了令人瞩目的成就。凭借着强大的表现能力和广泛的适用性, 盘古大模型在众多的应用场景中都表现出了不俗的表现。但在实际任务中如何有效地应用盘古大模型, 还有待于深入研究与实践。论文旨在对盘古大模型的技术原理及其在实际任务中的应用效果进行系统的探讨, 以期通过应用实训的方式, 为今后盘古大模型的应用方向提供参考。

## 2 盘古大模型概述

### 2.1 盘古大模型的基本架构

盘古大模型是一种基于 Transformer 架构的大规模预训练模型, 基本架构由多个编码器和解码器层组成, 每个层都包含自注意力机制和前馈神经网络。

盘古大模型采用基于 Transformer 架构的设计, 其编码

器与解码器部分均由多层模块堆叠组成。编码器的每一层包含两个主要子层: 多头自注意力机制和前馈神经网络。多头自注意力机制通过多个注意力头并行处理输入序列, 能够捕捉长距离依赖, 提升模型的特征表示能力。前馈神经网络则由两层全连接网络构成, 主要用于非线性变换, 进一步增强模型的表达能力。解码器部分同样由多个重复层组成, 但每层增加了一个编码器-解码器注意力机制。在此结构中, 多头自注意力机制用于解析解码器内部的依赖关系, 而编码器-解码器注意力机制则使解码器能够更好地利用编码器的输出, 为生成序列提供指导。前馈神经网络在解码器中的作用与编码器一致, 帮助增强模型的非线性处理能力。模型的输入首先经过嵌入层转化为固定维度的向量, 再依次通过编码器和解码器处理。在预训练阶段, 模型使用掩码语言模型(MLM)和下一句预测(NSP)任务, 以捕捉通用的语言特征<sup>[1]</sup>。在微调阶段, 盘古大模型借助特定任务的数据进行训练, 使其适应各种下游应用场景, 展现出较高的泛化能力。

### 2.2 盘古大模型的技术特点

盘古大模型具备多项关键技术特点, 使其在自然语言处理领域表现出色: ①大规模预训练, 通过在海量无监督语

**【作者简介】** 颜海鹰(2002-), 男, 中国湖南株洲人, 本科, 工程师, 从事人工智能研究。

料库上进行预训练，盘古大模型能够学习到丰富的语言表示，具备强大的泛化能力；②多模态融合，盘古大模型不仅支持文本数据，还能处理图像、音频等多种模态数据，实现跨模态任务的高效处理；③跨领域迁移能力，通过预训练学习到的通用表示，盘古大模型可以在不同领域和任务中进行迁移，提高模型的适应性和鲁棒性；④高效的并行计算，利用分布式训练技术，盘古大模型能够在大规模集群上高效训练，显著缩短训练时间；⑤模型压缩与加速，通过模型剪枝、知识蒸馏和量化技术，盘古大模型在保持性能的同时，显著减少参数数量和计算复杂度，提高推理速度。

### 2.3 盘古大模型的应用场景

盘古大模型在多个自然语言处理（NLP）应用场景中展现出卓越的性能：①文本分类，盘古大模型可以自动识别和分类文本内容，如新闻分类、垃圾邮件过滤等；②情感分析，通过分析文本中的情感倾向，盘古大模型能够判断文本的情感极性，如正面、负面或中立，广泛应用于社交媒体监控和市场调研；③问答系统，盘古大模型能够生成高质量的回答，提供智能客服和知识检索服务，提升用户体验；④机器翻译，盘古大模型支持多种语言对的高质量翻译，促进跨语言交流；⑤对话系统，通过生成连贯、自然的对话，盘古大模型能够提供人机交互服务，应用于虚拟助手和聊天机器人。

## 3 盘古大模型的技术原理

### 3.1 预训练与微调

盘古大模型的核心技术围绕预训练和微调展开。在预训练阶段，模型基于大规模无监督语料进行训练，以掌握语言的多样性和表达能力。常见的预训练任务包括掩码语言模型（MLM）和下一句预测（NSP）。MLM通过随机遮蔽输入序列中的部分单词，让模型预测被遮蔽的词语，从而学习上下文之间的依赖关系。NSP则通过判断两个句子是否连续，帮助模型理解句子间的逻辑关联。这些任务使模型能够在广泛的语境中建立起语言理解能力。完成预训练后，模型需要针对具体应用场景进行微调。微调阶段使用特定任务的有监督数据，通过调整模型参数使其更好地适应下游任务<sup>[2]</sup>。此过程采用与任务相关的损失函数，并借助反向传播算法对模型进行更新。预训练与微调的有机结合，使模型在各类实际应用中展现出卓越的性能与更好的泛化能力。公式如下：

$$Loss_{MLM} = -\frac{1}{|M|} \sum_{i \in M} \log P(w_i | w_{masked})$$

其中， $|M|$  为被遮蔽的单词数量； $w_i$  为第  $i$  个被遮蔽的单词； $w_{masked}$  为经过遮蔽处理的输入序列。

### 3.2 注意力机制

注意力机制是盘古大模型中的一个关键部件，主要分为自注意力机制和多头注意力机制两种。通过计算各位置与其他位置在输入序列中的相关性，自注意力机制会产生加权表示法。这种方式使模型不需要对所有信息进行均匀处理，而只关注输入序列中的重要部分。而多头注意力机制则是通

过并行的多个自我注意力机制来捕捉不同特征的输入序列。各头各自独立计算注意力，再通过线性转换层将结果拼接整合。多头关注机制可以从多个角度对输入序列进行关注，使模型增强了模型的表达能力。

### 3.3 损失函数与优化算法

在盘古大模型的训练过程中，损失函数和优化算法是确保模型性能的关键因素。损失函数用于衡量模型预测值与真实值之间的差异，常见的损失函数包括交叉熵损失（Cross-Entropy Loss）和均方误差损失（Mean Squared Error, MSE）。在预训练阶段，盘古大模型通常使用掩码语言模型（MLM）和下一句预测（NSP）任务的损失函数。例如，MLM的损失函数可以表示为：

$$Loss_{MLM} = -\frac{1}{|M|} \sum_{i \in M} \log P(w_i | w_{masked})$$

其中， $|M|$  表示被遮蔽的单词数量， $w_i$  是  $i$  个被遮蔽的单词， $w_{masked}$  是经过遮蔽处理的输入序列。在微调阶段，根据具体任务选择合适的损失函数。例如，对于分类任务，通常使用交叉熵损失：

$$Loss_{ce} = -\frac{1}{N} \sum_{i=1}^N y_i \log(p_i)$$

其中， $N$  是样本数量， $y_i$  是真实标签， $p_i$  是模型预测的概率。优化算法用于更新模型参数，以最小化损失函数。常用的优化算法包括 Adam、SGD（随机梯度下降）和 Adagrad。Adam 优化算法结合了动量和 RMSProp 的优点，能够有效地处理稀疏梯度和非平稳目标函数，更新规则如下：

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$$

其中， $m_t$  和  $v_t$  分别是梯度的一阶矩估计和二阶矩估计， $g_t$  是当前梯度， $\alpha$  是学习率， $\beta_1$  和  $\beta_2$  是衰减率， $\epsilon$  是一个小常数，用于防止除零错误。通过这些损失函数和优化算法，盘古大模型能够在大规模数据上高效训练，提升模型的性能和泛化能力<sup>[3]</sup>。

### 3.4 模型压缩与加速

在实际应用中，盘古大模型虽然性能强大，但其庞大的参数数量和计算复杂度往往成为部署和推理的瓶颈。通过模型压缩与加速技术，能够在保持模型性能的同时，显著减少模型的参数数量和计算资源消耗。

模型剪枝通过移除模型中不重要的权重或神经元，减少了参数，同时也减少了计算复杂度。剪枝方法可分为结构化剪枝和非结构化剪枝两种。结构化剪枝主要是对卷积层和

全联接层进行全通道或全线条的剪枝,使模型保持结构上的完整性。非结构化剪枝是以单重剪枝,对剪枝的粒度控制得比较精细。通过迭代剪枝和微调,大幅度缩小模型的尺寸和计算量,而不会对模型的性能造成重大影响。

知识蒸馏是将大模型的知识转移到小模型的一种技术。通过培养较小规模的学生模型来模仿大型教师模型的输出,在保持较小规模的同时,学生模型能够继承大部分教师模型的性能。具体做法是,不直接使用真正的硬标签,而使用教师模型的软标签作为学生模型的训练目标(即概率分布)。

量化技术通过将模型的权重和激活值从浮点数(如8位整数)转换成低精度的整数,减少了模型的存储和计算需求。量化有静态定量和动态定量之分。训练后进行静态量化,把模型的权重和激活值通过离线计算量化参数转化为低精度的表示方式<sup>[4]</sup>。动态量化则是在推理过程中动态量化,适用于需要实时推理的场景。量化技术不仅能减少模型的存储空间,还能加速推理过程。

## 4 盘古大模型应用实训

### 4.1 实训环境搭建

在进行盘古大模型的应用实训前,需搭建合适的环境。选择高性能计算资源,如配备 NVIDIA Tesla V100 和 A100 GPU 的服务器。安装 Python 3.8 及以上版本、PyTorch 1.7 及以上版本和 Transformers 库,确保所有依赖项正确安装,建议使用 Anaconda 环境管理工具。收集和整理适用于不同任务的数据集,如 IMDB 电影评论数据集(文本分类)、Twitter 情感分析数据集(情感分析)和 SQuAD 数据集(问答系统)。配置数据路径、模型保存路径和日志路径,确保文件和目录权限正确。编写配置文件,设置超参数和训练参数,如学习率、批量大小和训练轮次。

### 4.2 实训任务设计

为了验证盘古大模型的性能,设计了以下三个任务:

①文本分类,使用 IMDB 电影评论数据集,训练模型对电影评论进行正面或负面分类。评估指标包括准确率、精确率、召回率和 F1 分数。

②情感分析,使用 Twitter 情感分析数据集,训练模型识别推文的情感倾向。评估指标同上。

③问答系统,使用 SQuAD 数据集,训练模型回答阅读理解问题。评估指标包括准确率和 F1 分数。

### 4.3 实训过程与步骤

在盘古大模型的应用实训中,对分词、去除停词、建立词汇表等数据集进行清洗和预处理,以保证符合模型输入要求的数据格式。盘古大模型预训练加载完毕,按任务需要微调。学习率、批量大小、训练轮次等超参数的配置车型。在事先准备好的训练数据集中,会有一个模型的训练。训练时使用适当的损失函数(例如交叉熵损失)和优化算法(例如 Adam)。使用准确度、精确率、召回率和 F1 积分等指标来评价验证集合上的模型性能。车型参数根据评测结果进

行调整并进一步优化<sup>[5]</sup>。在测试数据集上做最后的测试,对没有看到的数据中模型的性能进行评估。对模型的性能指标进行分析,找出常见的错误并加以改进。通过具体案例,展示模型在实际应用中的表现和优点。

### 4.4 实训结果分析

盘古大模型在三个任务中均表现出色,具体如表1所示。

表1 实训结果

任务	指标	训练集	验证集	测试集
文本分类	准确率	95.2%	94.8%	94.5%
	精确率	95.5%	95.0%	94.7%
	召回率	95.0%	94.6%	94.3%
	F1 分数	95.2%	94.8%	94.5%
情感分析	准确率	94.5%	94.0%	93.8%
	精确率	94.8%	94.3%	94.0%
	召回率	94.2%	93.7%	93.5%
	F1 分数	94.5%	94.0%	93.8%
问答系统	准确率	88.0%	87.5%	87.2%
	F1 分数	88.5%	88.0%	87.7%

在文本分类任务中,模型的准确率、精确率、召回率和 F1 分数在测试集上分别达到了 94.5%、94.7%、94.3% 和 94.5%。情感分析任务中,各项指标在测试集上分别为 93.8%、94.0%、93.5% 和 93.8%。问答系统任务中,准确率和 F1 分数在测试集上分别为 87.2% 和 87.7%。这些结果显示,盘古大模型在不同任务中具有较高的性能和稳定性,能够在实际应用中提供可靠的解决方案。

## 5 结语

盘古大模型的应用实训为学习者提供了深入了解大规模预训练模型技术的机会,并通过实践掌握了模型的训练、微调和优化过程。通过搭建实训环境、设计任务和分析结果,使参与者既对自然语言加工技术有了认识上的促进作用,又将实际运用的经验积累了起来。该实训既促进了 AI 模型运用,又培养了参与者的创新能力和解决问题的能力,同时为以后的技术研发及工程实务奠定了扎实的基础。

### 参考文献

- [1] 吴锦浩,朱权洁,廖忠友,等.基于盘古大模型的矿用钢丝绳表面损伤检测研究[J].工业控制计算机,2024,37(1):1-3+6.
- [2] 孙思涵.大模型及其应用前景分析——从大模型应用场景到推理算力在边缘的展望[J].江西通信科技,2024(3):1-2.
- [3] 吕文哲,贾晨,唐华云.金融领域大语言模型应用安全探析[J].债券,2024(8):77-81.
- [4] 田辉,马茜,郑熠.大模型行业应用设计与实践[J].计算机与网络,2024,50(4):331-338.
- [5] 胡正坤,谭覃.生成式AI的大视听产业应用和未来展望——以商汤“日日新SenseNova”大模型为例[J].视听界,2024(4):27-29+38.