

CodeGeeX指导手册

1. 介绍

CodeGeeX 是一个由智谱AI研发的强大的智能编程助手，能够理解并生成各种编程语言的代码。它基于大型预训练语言模型，通过强大的生成能力来辅助开发人员完成代码编写任务。

CodeGeeX可完成的功能有：

- **代码自动生成和补全**：CodeGeeX可以根据自然语言注释描述的功能自动生成代码，也可以根据已有的代码自动生成后续代码，补全当前行或生成后续若干行，帮助你提高编程效率。
- **代码翻译**：基于AI大模型对代码进行语义级翻译，支持多种编程语言互译。
- **自动添加注释**：CodeGeeX可以给代码自动添加行级注释，节省大量开发时间。没有注释的历史代码，也不再是问题。
- **智能问答**：开发中遇到的技术问题，可直接向AI提问。无需离开IDE环境，去搜索引擎寻找答案，让开发者更专注地沉浸于开发环境。

本手册将指导你如何使用 MindNLP 框架来部署和使用 CodeGeeX 进行推理。

2. 安装 MindNLP

设置pip源为清华源，或者其他任何一个可用的源

```
pip config set global.index-url https://pypi.tuna.tsinghua.edu.cn/simple
```

安装mindspore和mindnlp等包：

```
%%capture captured_output
# 实验环境已经预装了mindspore==2.3.0，如需更换mindspore版本，可更改下面
MINDSPORE_VERSION 变量
!pip uninstall mindspore -y
!export MINDSPORE_VERSION=2.3.1
!pip install https://ms-release.obs.cn-north-4.myhuaweicloud.com/${MINDSPORE_VERSION}/MindSpore/unified/aarch64/mindspore-${MINDSPORE_VERSION}-cp39-cp39-linux_aarch64.whl --trusted-host ms-release.obs.cn-north-4.myhuaweicloud.com -i https://pypi.mirrors.ustc.edu.cn/simple
```

3. 加载和配置模型

3.1 导入必要的库

在你的 Python 脚本或 Jupyter Notebook 中，导入必要的库：

```
import mindspore
from mindnlp.core import no_grad
from mindnlp.transformers import AutoModelForCausalLM, AutoTokenizer
```

3.2 加载 Tokenizer 和模型

使用以下代码加载 CodeGeeX 的 tokenizer 和模型：

```
# 加载tokenizer文件
tokenizer = AutoTokenizer.from_pretrained("THUDM/codegeex2-6b",
mirror='huggingface')

# 加载模型
model = AutoModelForCausalLM.from_pretrained(
    "THUDM/codegeex2-6b",
    mirror='huggingface',
    ms_dtype=mindspore.float16,
).eval()
```

4. 使用模型进行推理

4.1 编码输入

定义你想要与 CodeGeeX 对话的内容，并对其进行编码：

```
# 在此输入希望和CodeGeeX对话的内容
query = "write me a bubble sort algorithm in python"

inputs = tokenizer(query, return_tensors="ms")
print(inputs)
```

4.2 生成输出

设置推理参数并生成代码：

```
# 设置采样、最大生成长度等配置，开始推理
# _framework_profiler_step_start()
gen_kwargs = {"max_length": 100, "do_sample": True, "top_k": 1}
with no_grad():
    outputs = model.generate(**inputs, **gen_kwargs)
    outputs = outputs[:, inputs['input_ids'].shape[1]:]
    print(tokenizer.decode(outputs[0], skip_special_tokens=True))
# _framework_profiler_step_end()
```

5. 注意事项

- 确保你已经安装了 MindNLP 和 MindSpore 的正确版本。
- 根据你的实际需求调整推理参数，如 `max_length` 和 `top_k`。
- 如果遇到模型配置相关的错误，检查配置文件或尝试其他模型。

6. 参考文献

- [MindNLP 官方文档](#)
- Hugging Face Model Hub
- MindSpore 官方文档