Android

Build on Android using Termux（使用Termux在Android进行构建）

Termux is a method to execute `llama.cpp` on an Android device (no root required).（Termux是一种可以在Android设备上运行llama.cpp的方法。）

```
# 更新包管理器索引
apt update && apt upgrade -y
# 安装git、make、cmake
apt install git make cmake
```

It's recommended to move your model inside the `~/` directory for best performance（为了最佳的性能将模型移动到~/目录中是非常推荐的）:

```
# 将当前工作目录切换到storage/downloads中
cd storage/downloads
# 将model.gguf模型文件移动到~/文件目录中
mv model.gguf ~/
```

Get the code & follow the Linux build instructions to build `llama.cpp`.

- 根据给出的guide对llama.cpp进行构建

Building the Project using Android NDK（使用Android NDK构建项目）

Obtain the Android NDK and then build with CMake.（获取Android NDX并且使用CMake构建项目）

Execute the following commands on your computer to avoid downloading the NDK to your mobile. Alternatively, you can also do this in Termux（在你的电脑上执行下列的命令从而避免在你的移动设备上下载NDX。或者你也可以使用Termux做这件事）:

```
$ mkdir build-android
$ cd build-android
$ export NDK=<your_ndk_directory>
$ cmake -DCMAKE_TOOLCHAIN_FILE=$NDK/build/cmake/android.toolchain.cmake -
DANDROID_ABI=arm64-v8a -DANDROID_PLATFORM=android-23 -DCMAKE_C_FLAGS=-
march=armv8.4a+dotprod ..
$ make
```

Install termux on your device and run `termux-setup-storage` to get access to your SD card (if Android 11+ then run the command twice).

- 在你的设备上安装termux并且通过运行termux-setup-storage命令获取SD卡的访问权限。

Finally, copy these built `llama` binaries and the model file to your device storage. Because the file permissions in the Android sd card cannot be changed, you can copy the executable files to the `/data/data/com.termux/files/home/bin` path, and then execute the following commands in Termux to add executable permission:

- 最后将这些构建好的llama二进制文件和模型文件拷贝到你的设备存储器上。因为在Android sd card的文件权限不能被更改，你可以将可执行文件拷贝到/data/data/com.termux/files/home/bin路径中，然后在termux中执行下列命令从而添加可执行权限。

(Assumed that you have pushed the built executable files to the /sdcard/llama.cpp/bin path using `adb push`)
（假设你已经通过adb push将构建了的可执行文件放到了/sdcard/llama.cpp/bin路径下）

```
$cp -r /sdcard/llama.cpp/bin /data/data/com.termux/files/home/
$cd /data/data/com.termux/files/home/bin
$chmod +x ./*
```

Download model llama-2-7b-chat.Q4_K_M.gguf, and push it to `/sdcard/llama.cpp/`, then move it to `/data/data/com.termux/files/home/model/`

- 下载llama-2-7b-chat.Q4_K_M.gguf模型，然后将这个模型文件push到/sdcard/llama.cpp/中，然后将其移动到/data/data/com.termux/files/home/model/中。

```
$mv /sdcard/llama.cpp/llama-2-7b-chat.Q4_K_M.gguf
/data/data/com.termux/files/home/model/
```

Now, you can start chatting:

```
$cd /data/data/com.termux/files/home/bin
$./llama-cli -m ../model/llama-2-7b-chat.Q4_K_M.gguf -n 128 -cml
```

Here's a demo of an interactive session running on Pixel 5 phone:

https://user-images.githubusercontent.com/271616/225014776-1d567049-ad71-4ef2-b050-55b0b3b9274c.mp4