

使用GNU Parallel进行FreeSurfer并行计算

Alex / 2024-07-30 / free_learner@163.com / AlexBrain.cn

一、背景

一个项目通常会有上百个被试的数据，一个被试数据跑完FreeSurfer的recon-all流程通常需要数个小时（主要取决于FreeSurfer的版本和CPU的性能），所以如果是逐个被试顺序执行，那么整个项目需要数周的时间。通过并行计算，可以极大缩短项目周期。并行计算可以分为被试内的并行计算和被试间的并行计算。被试内的并行计算是指在一个被试的recon-all流程中，有些步骤可以并行计算；被试间的并行计算是指同时运行多个被试的recon-all流程。以下测试中，我的电脑CPU是AMD Ryzen threadripper pro 7965wx，在VirtualBox的虚拟机中进行计算，操作系统是Ubuntu 22.04 LTS。

二、被试内并行计算

通常情况下，我们可以使用如下命令进行recon-all流程，这种情况下没有任何并行计算：

```
recon-all -sd ${OUTDIR} -i ${INDIR}/S01.nii.gz -s S01 -all
```

在上面的命令中，`${INDIR}/S01.nii.gz` 表示一个被试的T1数据，`${OUTDIR}` 表示存放结果的目录，`S01` 表示被试结果文件夹的名字，即最后的结果存放在 `${OUTDIR}/S01` 目录下。

FreeSurfer的recon-all命令提供了两个选项用于并行计算：`-parallel` 和 `-openmp num`（实际上，还有一个 `-threads num` 选项，该选项和 `-openmp num` 效果完全一样）。根据FreeSurfer的说明，`-parallel` 选项是两个半球的计算分开进行，而 `-openmp num` 表示指定 `num` 个进程进行计算。这两个选项也可以结合起来使用。不过官方的说明并没有很清楚，我前面的表述可能有误。具体命令如下：

```
## Using -parallel
recon-all -sd ${OUTDIR} -i ${INDIR}/S01.nii.gz -s S01 -all -parallel
## Using -openmp
recon-all -sd ${OUTDIR} -i ${INDIR}/S01.nii.gz -s S01 -all -openmp 4
```

在我电脑上的测试结果如下：

Default	-parallel	-openmp 4
2.72 hours	1.99 hours	1.78 hours

可以看到在使用并行计算选项后，运行时间减少了大约30%。注意这只是一次运行的结果，不太稳定。

三、使用GNU Parallel进行被试间并行计算

GNU Parallel是一个用于并行执行计算任务的工具，使用简单，适合在个人电脑上使用。

1. 安装GNU Parallel

```
sudo apt-get install parallel
```

通过包管理器安装的GNU Parallel不是最新版，如果需要最新版，需要去官网下载源码进行编译。

2. 基本用法

使用如下命令进行被试间的并行计算：

```
## Specify the subjects directly
parallel --jobs 2 --joblog ${OUTDIR}/mylog.txt recon-all -sd ${OUTDIR} -i ${OUTDIR}/{1}.nii.gz
-s {1} -all ::: S01 S02
## Specify the subjects from a file
parallel --jobs 4 --joblog ${OUTDIR}/mylog.txt recon-all -sd ${OUTDIR} -i ${OUTDIR}/{1}.nii.gz
-s {1} -all ::: $(cat sublist.txt)
```

在上面的命令中，`--jobs 2` 表示同时进行两个计算任务，`--joblog` 表示生成日志信息，用于查看哪些任务已经结束，`:::` 表示需要并行的参数信息，`{1}` 表示参数的占位符。GNU Parallel功能很丰富，我这里无法一一列举，也没有必要，因为网上相关文档和教程很多。

在运行之前，可以加上 `--dry-run` 参数，它会列出实际运行的命令（而不会执行），这样可以检查设置是否有误，比如对于输入参数的解析是否正确。例如：

```
parallel --dry-run --jobs 2 --joblog ${OUTDIR}/mylog.txt recon-all -sd ${OUTDIR} -i
${OUTDIR}/{1}.nii.gz -s {1} -all ::: S01 S02
```