Intergation of Openfold into MSI

Jack Raymakers

University of Minnesota

May 3, 2025

Table of Contents

Objective and Scope

Create an environment for the openfold project from the aqlaboratory/openfold repository, implement using the GPUs on MSI for Jupiter infrastructure and execute unit test and create a prediction using the openfold model. Document all steps.

Required Environment

Create a ticket

- Request a CSE Labs account,
  if you don't already have one
- Request a MSI services
- Optional Virtual Online Linux Environment (VOLE)

URL to Create ticket

`https://tdx.umn.edu/TDClient/31/Portal/Requests/`
`ServiceDet?ID=237`

- Install AnyConnect VPN Client Software(per os)
  `https://software.cisco.com/download/home/`
  `286330811/type/282364313/release/5.1.8.122?i=!pp`

open terminal session using vpn

- Launch vpn (open
  AnyConnect-UofMvpnFull,username,passward)
  $/opt/cisco/secureclient/bin/vpn or (vpnui for gui)
  VPNtc-vpn-1.vpn.umn.edu
  Group: [AnyConnect-UofMvpnFull]
  username:$USER
  Password:$secret

- open session to connect to MSI (ssh exmaple)
  ssh username@agate.msi.umn.edu
  Password:

Make Project path and clone repos

- Create Project Workspace and change directory
  example mkdir -p /scratch.global/$USER/git
  example cd /scratch.global/$USER/git/
- full path for example **fullpath** is /scratch.global/$USER/git
- clone repo's
  git clone –recursive https://github.com/NVIDIA/dllogger.git
  git clone –recursive
  https://github.com/Dao-AILab/flash-attention
  git clone https://github.com/NVIDIA/cutlass –depth 1
- clone openfold from custimize repo
  git clone https://github.com/rayma017/openfold
- The start point was: aqlaboratory/openfold -b pl_updates

Use mamba to create openfold-env

- customize .condarc to save space on home drive
  conda config –set auto_activate_base false
  conda config –add envs_dirs
  /scratch.global/$USER/.conda/envs_dir
  conda config –add pkgs_dirs
  /scratch.global/$USER/.conda/pkgs_dir
- run mamba –cd openfold
  ./environment.yml.sh
- if you have a completion issue navigate to project dir use
  mamba with update
  mamba env update -f environment.yml
- after successful creation you be provide with activate sequence

Quick Check

- Hop to GPU node
  srun -N 1 -n 1 –mem=32g -t 240 –gres=gpu:a40:1
  –tmp=100g -p interactive-gpu –x11 xterm -fa default -fs 1
- Run... conda deactivate
- Run... conda activate openfold-env
- Run... python test_cuda.py

copy model parameters to your update openfold/resources

- Navigate to inside repo and execute
- update conda environmet source scripts/update_conda_vars
- scripts/download_openfold_params.sh resources/
- scripts/download_openfold_solseq_params.sh resources/
- scripts/download_alphafold _params.sh resources/

- Navigate to inside openfold repo exec:
- run unit test
  srun -N 1 -n 10 –mem=32g -t 100 –gres=gpu:a40:1 \
  –tmp=100g -p interactive-gpu –pty bash
  conda actavate openfold-env scripts/run_unit_tests.sh
- Make prediction using example
  cd tests/test_data # Please note the main database is not
  install
  ../../examples/monomer/inference.sh

Current state and Notes

- openfold model, openfold parameters, alphafold parameters,utilizes and tests are installed
- Several Terabytes of date from protein database are install
- Several Terabytes of training data are install
- ran unit test
- ran prediction with model
- updated original environment to run on MSI
- created a git with updates
- TODO
- Create Jupiter note example on MSI
- Verify git/raym017/openfold with fresh install

Notes and Questions

- Resolving conflict cuda ,pytorch and requirement from other install
- Downloading data exceeded time allocation
- Urls that no longer existed
- GUI: launching plots html,raster
- Resolving help gain structure of the database
- Questions: