SDFT Day 1: AI Workflow Examples

Keywords: computer vision, image normalization, classifiers, NLP, speech, transformers models

Learning Outcomes

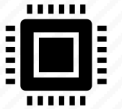
AI Workflow for Computer Vision

- Learn about sources of variability in image data and challenges in working with large images
- Gain familiarity with common steps in and image analysis pipeline and their computational requirements

NLP and Speech Pipelines

- Review classical ML techniques for NLP (Naïve Bayes, SVM, Linear Regression)
- Gain familiarity with Transformer-based model tuning (BERT)
- Understand commonalities and differences between pipelines for classical and transformer-based models

Exercises

Walkthrough of Newsgroup Classification pipeline with Classical ML		
Walkthrough of Newsgroup Classification pipeline with BERT finetuning		
Optional: Speech Recognition Pipeline – digit recognizer		

Files: /fs/ess/PZS1124/AI_BOOTCAMP_EF

SDFT Day 2: HPC Resources for AI and Python support at OSC

Keywords: HPC systems, GPUs/accelerators, data management, python, jupyter, conda, pip

Learning Outcomes

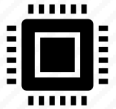
HPC Resources for AI

- Understand common HPC system makeup and organization
- Identify system components that are most critical for AI workflows
- Understand that supporting notebooks and data transfer tools are equally as important as hardware

Python at OSC

- Understand how python is integrated with software environment management (modules)
- Gain familiarity with common python package management tools and their use
- Consider support for python version and custom package defaults for python notebook deployments

Exercises

Explore use of the NIH National Cancer Institute (NCI)Genomic Data Commons (GDC) data transfer tool		
Practice installing python packages with conda and pip		
Create a custom jupyter kernel to use with interactive notebook		

/fs/ess/PZS1124/AI_BOOTCAMP_KT/su2023-day1/README.txt

SDFT Day 3: Advanced Parallelization Strategies for Distributed DNN Training

Keywords: data parallelism, model parallelism, distributed training, distributed inference, inference frameworks

Learning Outcomes

Parallelization Strategies

- Understand data parallelism approach and communication requirements for distributed training
- Get introduced to model layer level, neuron-level and hybrid parallelism techniques
- Be exposed to tradeoffs due to memory requirements for ML and choices for parallelization

Distributed Inferencing

- Gain familiarity with inference frameworks and supported parallelism

Exercises

Explore batch sizes for DDN training with multi-GPU with ResNet model and FashionMNIST dataset		
Compare training using pipeline parallelism for 'out-of-core' DNN models using Pytorch-Gpipe.		

/fs/ess/PZS1124/AI_BOOTCAMP_DK/session3/

SDFT Day 4: Advanced Parallelization Strategies for Distributed Inference and Hyperparameter Optimization

Keywords: DL Inference, quantization, Hyperparameter Optimization, DeepSpeed, DeepHyper

Learning Outcomes

Distributed Inference

- Understand how DL inference differs from training
- Be familiar with different inference scenarios: online vs batch, data center vs. edge
- Become aware of quantization for reducing model size

Distributed DNN Training/Inference using DeepSpeed

- Become aware of tools that optimize large models to make training more tractable

Scalable Hyperparameter Optimization using DeepHyper

- Understand the difference between a model parameter and a hyperparameter use to control learning
- Become familiar with hyperparameter optimization be aware of common optimization algorithms

Exercises

Explore training 2.5B parameter BERT model with DeepSpeed and ZeRO		
Explore using DeepHyper to optimize hyperparmeters for a text classification model		

AI Bootcamp Project, NSF AWD 2118250

SDFT Day 5: Tools for Understanding and Debugging

Keywords: profiling, debugging, Pytorch-Lightning, Tensorboard, visualizations for NNs

Learning Outcomes

Tools for debugging python code

- Become aware of best practices and available tools for debugging
- Become aware of capabilities of HPC debugging tools

Tools for understanding performance of python code

- Understand uses and general capabilities of software profiling tools

Tools for managing and understanding ML training

- Become aware of tools including Pytorch-Lightning and Tensorboard and their basic capabilities
- Learn about some visualizations techniques that are helpful for understanding training results.

Exercises

Explore using Linero DDT to debug a toy python example.		
Explore using the cPython package and Linero MAP to profile python code.		
Demonstration of Pytorch-Lightning and TensorBoard		

`/fs/ess/PZS1124/AI_BOOTCAMP_KT/su23-day5/`

SDFT Day 6: Advanced HPC Technology

Keywords: CPU Chips, GPUs, DPUs, AI Accelerators, Communication middleware, OneAPI

Learning Outcomes

CPUs

- Become familiar with current CPU technology from leading chip vendors: Intel, AMD and Arm

GPUs

- Understand how GPUs differ from CPU.
- Become familiar with current GPU technology from leading chip vendors: NVIDIA, AMD and Intel

DPUs and AI Accelerators

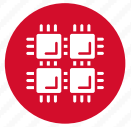
- Be exposed to network DPUs and AI accelerator devices

Software stacks

- Understand the possible integrations between MPI, GPU/Accelerator and network
- Be exposed to a comprehensive software stack; Intel's OneAPI as an example

Exercises

No exercises



SDFT Day 7: AI Accelerator Testbeds

Keywords: AI Accelerators, Cerebras, Habana, Graphcore, Kubernetes

Learning Outcomes

SDSC Voyager Testbed

- Learn about the Voyager project and getting system access
- Be exposed to the Intel Habana Gaudi and Inference devices and programming environment

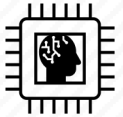
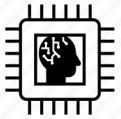
PSC Neocortex Resource

- Learn about the Neocortex project getting system access
- Be exposed to the Cerebras wafer capabilities and development environment

TAMU ACES Graphcore IPU

- Learn about the ACES Testbed and getting system access
- Be exposed to the Graphcore IPU capabilities and software stack

Exercises

Demo of running a TensorFlow example with Kubernetes on Voyager		
Train an image classifier on Neocortex using the MNIST data set.		
Convert a pytorch NN model to run on the IPU		

AI Bootcamp Project, NSF AWD 2118250

SDFT Day 8: Large Language Models

Keywords: LLMs, Transformers, encoder, decoder, instruction tuning, reinforcement learning, hallucinations

Learning Outcomes

LLM Development and scaling

- Review the transformer models and their evolution
- Understand what 'large' means in terms of parameters, training data and computing requirements
- Learn about commonly used models such as ChatGPT, PaLM, LLaMA, etc.

LLM Tuning techniques

- Be introduced to common model tuning approaches including instruction tuning and reinforcement learning

LLM capabilities

- Become familiar with LLM capabilities on common benchmark problems
- Consider opportunities for using LLMs given current strengths
- Be aware of some of the limitations of current models such as producing hallucinations

Exercises

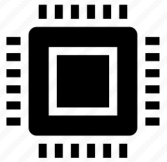
Submit a request for an LLM, discuss cases where it might produce hallucinations	n/a	n/a

AI Bootcamp Project, NSF AWD 2118250

Key to icons for exercises

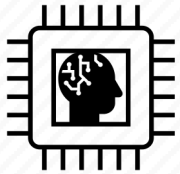
Command Line

Notebook



CPU

GPU



AI Accelerator