

So sánh PyTorch với PyTorch Lightning

Giáo viên hướng dẫn: Thầy Đặng Tuấn Linh

Sinh viên: Trần Văn Tài 20207700



GIỚI THIỆU:

PyTorch là một thư viện mã nguồn mở cho việc tính toán khoa học và học máy sử dụng ngôn ngữ Python. Pytorch được phát triển bởi Facebook AI Research và có mục tiêu cung cấp một nền tảng linh hoạt cho việc xây dựng các mô hình học máy.



CÁC GÓI TRONG PYTORCH

1.torch: gói chính của PyTorch và chứa các cấu trúc dữ liệu cơ bản như tensor, các hàm toán học và các phép toán trên tensor. Là cốt lõi của PyTorch và cung cấp nền tảng để thực hiện tính toán số học và xây dựng mô hình.

`tensor`

Constructs a tensor with no autograd history (also known as a "leaf tensor", see [Autograd mechanics](#)) by copying `data`.

`reshape`

Returns a tensor with the same data and number of elements as `input`, but with the specified shape.

Tìm hiểu thêm: <https://pytorch.org/docs/stable/torch.html>



CÁC GÓI TRONG PYTORCH

2.torch.nn: chứa các lớp và hàm hỗ trợ để xây dựng mạng nơ-ron, cung cấp các lớp như Conv2d, Linear, các hàm loss như CrossEntropyLoss, MSELoss và các hàm kích hoạt như ReLU, Sigmoid, và Softmax, giúp đơn giản hóa việc xây dựng và tùy chỉnh mô hình.

`nn.Conv2d`

Applies a 2D convolution over an input signal composed of several input planes.

`nn.ReLU`

Applies the rectified linear unit function element-wise:

Tìm hiểu thêm: <https://pytorch.org/docs/stable/nn.html>



CÁC GÓI TRONG PYTORCH

3.torch.optim: chứa các thuật toán tối ưu hóa cho huấn luyện mô hình, cung cấp các thuật toán như SGD, Adam, và RMSprop để điều chỉnh các tham số mô hình theo cách tốt nhất.

```
optimizer = optim.SGD(model.parameters(), lr=0.01, momentum=0.9)
optimizer = optim.Adam([var1, var2], lr=0.0001)
```

Tìm hiểu thêm: <https://pytorch.org/docs/stable/optim.html>



CÁC GÓI TRONG PYTORCH

4.torch.utils.data: cung cấp các công cụ để làm việc với dữ liệu và tạo batch dữ liệu cho huấn luyện, chứa các lớp như Dataset và DataLoader để tải và chia dữ liệu thành các batch.

```
Dataloader(dataset, batch_size=1, shuffle=False, sampler=None,  
            batch_sampler=None, num_workers=0, collate_fn=None,  
            pin_memory=False, drop_last=False, timeout=0,  
            worker_init_fn=None, *, prefetch_factor=2,  
            persistent_workers=False)
```

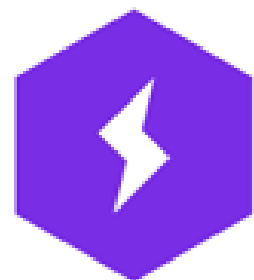
Tìm hiểu thêm: <https://pytorch.org/docs/stable/data.html>



CÁC GÓI TRONG PYTORCH

5.torchvision: cung cấp các công cụ và tập dữ liệu phổ biến cho thị giác máy tính, chứa các tập dữ liệu như MNIST, CIFAR-10, và ImageNet, cùng với các phép biến đổi dữ liệu và các mô hình thị giác như ResNet và AlexNet.

Tìm hiểu thêm: <https://pytorch.org/vision/stable/index.html>



PyTorch Lightning

GIỚI THIỆU:

PyTorch Lightning là một framework sử dụng trên nền tảng PyTorch, giúp đơn giản hóa và tăng cường hiệu suất của quy trình huấn luyện mô hình trong Machine Learning và Deep Learning.

PyTorch Lightning mở rộng PyTorch bằng cách cung cấp một kiến trúc và các tiện ích bổ sung để giảm thiểu việc lặp lại mã, quản lý chu kỳ huấn luyện, ghi nhật ký, phân tách phần giải quyết vấn đề và quy tắc tối ưu hóa.



PyTorch Lightning

CÁC GÓI TRONG PYTORCH LIGHTNING

1.pytorch_lightning: package chính của PyTorch Lightning, cung cấp một giao diện dễ sử dụng và mô hình chuẩn hóa để huấn luyện mô hình PyTorch.

Tìm hiểu thêm: <https://www.pytorchlightning.ai/index.html>



PyTorch Lightning

CÁC GÓI TRONG PYTORCH LIGHTNING

2.pytorch_lightning.callbacks: chứa các lớp callback được sử dụng trong quá trình huấn luyện, bao gồm ModelCheckpoint, EarlyStopping, LearningRateMonitor.

`EarlyStopping`

Monitor a metric and stop training when it stops improving.

`ModelCheckpoint`

Save the model periodically by monitoring a quantity.

Tìm hiểu thêm: <https://lightning.ai/docs/pytorch/stable/extensions/callbacks.html>



PyTorch Lightning

CÁC GÓI TRONG PYTORCH LIGHTNING

3.pytorch_lightning.logger: chứa các lớp logger để ghi lại thông tin quan trọng trong quá trình huấn luyện. Ví dụ: TensorBoardLogger, CSVLogger, NeptuneLogger.

`TensorBoardLogger`

Log to local file system in `TensorBoard` format.

```
from lightning.pytorch import loggers as pl_loggers

tb_logger = pl_loggers.TensorBoardLogger(save_dir="logs/")
trainer = Trainer(logger=tb_logger)
```

Tìm hiểu thêm: <https://lightning.ai/docs/pytorch/stable/extensions/logging.html>



PyTorch Lightning

CÁC GÓI TRONG PYTORCH LIGHTNING

4.pytorch_lightning.trainer: chứa các lớp Trainer để huấn luyện và kiểm tra mô hình PyTorch Lightning. Nó cung cấp nhiều tính năng như auto-scaling, distributed training.

```
model = MyLightningModule()

trainer = Trainer()
trainer.fit(model, train_dataloader, val_dataloader)
```

```
trainer.test(dataloaders=test_dataloaders)
```

Tìm hiểu thêm: <https://pytorch.org/docs/stable/nn.html>



PyTorch Lightning



SO SÁNH PYTORCH VÀ PYTORCH LIGHTNING

PyTorch Lightning là một framework cao cấp và tối ưu hóa cho PyTorch, giúp bạn giảm bớt công việc lặp lại và tập trung vào việc thiết kế và nghiên cứu mô hình học sâu,

PyTorch Lightning cung cấp một cấu trúc tổ chức rõ ràng, tích hợp các công nghệ phổ biến và hỗ trợ huấn luyện phân tán.

Về tốc độ: Không có sự khác biệt rõ ràng giữa PyTorch và PyTorch Lightning