Topics
My posts
More
CATEGORIES
General
Site Feedback
All categories

2025 USA-NA-AIO Round 2, Problem 3, Part 9

USAAIO May 2025

Part 9 (5 points, coding task)

Part 9.1

Define your own collate function.

- The function name is `my_collate_fn`.
- Padding
  - For text data, let the longest sample be with `K` tokens.
  - Consider another text sample with `L` tokens satisfying `L < K`. Then, in addition to those `L` tokens, this sample is padded with `K-L` padding tokens whose values are 0.
- Outputs
  - `token_id_batch`. If the batch size is `B` and the longest sample in the text data has `K` tokens, then `token_id_batch` is a tensor with shape `(B,K)`.
  - `attention_mask_batch`. This is a tensor that has shape `(B,K)`. If a position is occupied by a non-padding token, its value is 1. Otherwise, if it is occupied by a padding token, its value is 0. Data types are `int64`.
  - `image_batch`. This is a tensor that has shape `(B,3,224,224)`.

Part 9.2

Define a DataLoader object called `CLIP_dataloader`.

Skip to main content

- Set `batch_size = 16`.
- Set `shuffle = True`.
- Use the collate function defined in Part 10.

USAAIO **May 2025**

```
### WRITE YOUR SOLUTION HERE ###

# Part 9.1

def my_collate_fn(batch):
    image_batch_input, token_id_batch_input = zip(*batch)

    image_batch = torch.stack(image_batch_input)

    max_len_token_id = max([len(token_id) for token_id in token_id_batch_input])
    token_id_batch = []
    attention_mask_batch = []

    for token_id in token_id_batch_input:
        token_id_batch.append(torch.concatenate([token_id, torch.zeros(max_len_to
        attention_mask_batch.append(torch.concatenate([torch.ones(len(token_id),
                                                                  torch.zeros(max_len_token_

    token_id_batch = torch.stack(token_id_batch)
    attention_mask_batch = torch.stack(attention_mask_batch)

    return image_batch, token_id_batch, attention_mask_batch


# Part 9.2

batch_size = 16
```

Skip to main content

```
CLIP_dataloader = DataLoader(CLIP_dateset, batch_size = batch_size, shuffle = Tru

""" END OF THIS PART """
```

Related topics

Topic	Replies	Activity
2025 USA-NA-AIO Round 2, Problem 3, Part 8	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 6	1	May 2025
2025 USA-NA-AIO Round 2, Problem 1, Part 5	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 11	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 4	1	May 2025