

LLM Course

Search documentation

EN

3,022

0. SETUP

1. TRANSFORMER MODELS

2. USING TRANSFORMERS

3. FINE-TUNING A PRETRAINED MODEL

Introduction

Processing the data

Fine-tuning a model with the Trainer API or Keras

A full training

Fine-tuning, Check!

End-of-chapter quiz

4. SHARING MODELS AND TOKENIZERS

5. THE DATASETS LIBRARY

6. THE TOKENIZERS LIBRARY

7. CLASSICAL NLP TASKS

8. HOW TO ASK FOR HELP

9. BUILDING AND SHARING DEMOS

10. CURATE HIGH-QUALITY DATASETS

11. FINE-TUNE LARGE LANGUAGE MODELS

12. BUILD REASONING MODELS

COURSE EVENTS

PyTorch TensorFlow

Ask a question

End-of-chapter quiz

Test what you learned in this chapter!

1. The emotion dataset contains Twitter messages labeled with emotions. Search for it in the Hub , and read the dataset card. Which of these is not one of its basic emotions?

Joy

Love

Confusion

Surprise

Correct! Correct! Confusion is not one of the six basic emotions.

Submit You got all the answers!

2. Search for the ar\_sarcasm dataset in the Hub . Which task does it support?

Sentiment classification

Machine translation

Named entity recognition

Question answering

Correct! That's right! You can tell thanks to the tags.

Submit You got all the answers!

Profile kleberga

Notifications

Inbox (0)

End-of-chapter quiz

1. TI

2. Twit

3. emc

4. and

5. thes

6. emc

7. Create organization

8. Usage Quota

9. Private Storage

10. Zero GPU

11. Inference Usage

12. Get Hugging Face PRO

13. Settings

14. Access Tokens

15. Billing

16. Changelog

17. Sign Out

18. What's the purpose of Training Arguments ?

19. Why should you use the Accelerate library?

20. What happens when you instantiate one of the TFAutoModelForSeq2SeqWrapper classes with a pretrained language model (such as bert-base-uncased ) that corresponds to a

LLM Course

Search documentation

EN

3,022

0. SETUP

1. TRANSFORMER MODELS

2. USING TRANSFORMERS

3. FINE-TUNING A PRETRAINED MODEL

Introduction

Processing the data

Fine-tuning a model with the Trainer API or Keras

A full training

Fine-tuning, Check!

End-of-chapter quiz

4. SHARING MODELS AND TOKENIZERS

5. THE DATASETS LIBRARY

6. THE TOKENIZERS LIBRARY

7. CLASSICAL NLP TASKS

8. HOW TO ASK FOR HELP

9. BUILDING AND SHARING DEMOS

10. CURATE HIGH-QUALITY DATASETS

11. FINE-TUNE LARGE LANGUAGE MODELS

12. BUILD REASONING MODELS

COURSE EVENTS

3. How does the BERT model expect a pair of sentences to be processed?

Tokens\_of\_sentence\_1 [SEP] Tokens\_of\_sentence\_2

[CLS] Tokens\_of\_sentence\_1 Tokens\_of\_sentence\_2

[CLS] Tokens\_of\_sentence\_1 [SEP] Tokens\_of\_sentence\_2 [SEP]

[CLS] Tokens\_of\_sentence\_1 [SEP] Tokens\_of\_sentence\_2

Correct! That's correct!

Submit You got all the answers!

4. What are the benefits of the Dataset.map() method?

The results of the function are cached, so it won't take any time if we re-execute the code.

It can apply multiprocessing to go faster than applying the function on each element of the dataset.

It does not load the whole dataset into memory, saving the results as soon as one element is processed.

Correct! That's indeed one of the neat benefits of this method! It's not the only one, though...

Correct! This is a neat feature of this method, but it's not the only one!

Correct! That's one advantage of this method. There are others, though!

Submit You got all the answers!

5. What does dynamic padding mean?

It's when you pad the inputs for each batch to the maximum length in the whole dataset.

It's when you pad your inputs when the batch is created, to the maximum length of the sentences inside that batch.

It's when you pad your inputs so that each sentence has the same number of tokens as the previous one in the dataset.

Correct! That's correct! The "dynamic" part comes from the fact that the size of each batch is determined at the time of creation, and all your batches might have different shapes as a result.

Submit You got all the answers!

End-of-chapter quiz

1. The emotion dataset contains Twitter messages labeled with emotions. Search for it in the Hub , and read the dataset card. Which of these is not one of its basic emotions?

2. Search for the ar\_sarcasm dataset in the Hub . Which task does it support?

3. How does the BERT model expect a pair of sentences to be processed?

4. What are the benefits of the Dataset.map() method?

5. What does dynamic padding mean?

6. What is the purpose of a collate function?

7. What happens when you instantiate one of the TFAutoModelForSeq2SeqWrapper classes with a pretrained language model (such as bert-base-uncased ) that corresponds to a different task than the one for which it was trained?

8. What's the purpose of Training Arguments ?

9. Why should you use the Accelerate library?

10. What happens when you instantiate one of the TFAutoModelForSeq2SeqWrapper classes with a pretrained language model (such as bert-base-uncased ) that corresponds to a different task than the one for which it was trained?

11. The TensorFlow models from

LLM Course

Search documentation

EN

3,022

0. SETUP

1. TRANSFORMER MODELS

2. USING TRANSFORMERS

3. FINE-TUNING A PRETRAINED MODEL

Introduction

Processing the data

Fine-tuning a model with the Trainer API or Keras

A full training

Fine-tuning, Check!

End-of-chapter quiz

4. SHARING MODELS AND TOKENIZERS

5. THE DATASETS LIBRARY

6. THE TOKENIZERS LIBRARY

7. CLASSICAL NLP TASKS

8. HOW TO ASK FOR HELP

9. BUILDING AND SHARING DEMOS

10. CURATE HIGH-QUALITY DATASETS

11. FINE-TUNE LARGE LANGUAGE MODELS

12. BUILD REASONING MODELS

COURSE EVENTS

5. What does dynamic padding mean?

☐ It's when you pad the inputs for each batch to the maximum length in the whole dataset.

☒ It's when you pad your inputs when the batch is created, to the maximum length of the sentences inside that batch.

Correct! That's correct! The "dynamic" part comes from the fact that the size of each batch is determined at the time of creation, and all your batches might have different shapes as a result.

☐ It's when you pad your inputs so that each sentence has the same number of tokens as the previous one in the dataset.

Submit

You got all the answers!

6. What is the purpose of a collate function?

☐ It ensures all the sequences in the dataset have the same length.

☒ It puts together all the samples in a batch.

Correct! Correct! You can pass the collate function as an argument of a `DataLoader`. We used the `DataCollatorWithPadding` function, which pads all items in a batch so they have the same length.

☐ It preprocesses the whole dataset.

☐ It truncates the sequences in the dataset.

Submit

You got all the answers!

7. What happens when you instantiate one of the `AutoModelForXxx` classes with a pretrained language model (such as `bert-base-uncased`) that corresponds to a different task than the one for which it was trained?

☐ Nothing, but you get a warning.

☒ The head of the pretrained model is discarded and a new head suitable for the task is inserted instead.

Correct! Correct. For example, when we used `AutoModelForSequenceClassification` with `bert-base-uncased`, we got warnings when instantiating the model. The pretrained head is not used for the sequence classification task, so it's discarded and a new head is instantiated with random weights.

☐ The head of the pretrained model is discarded.

☐ Nothing, since the model can still be fine-tuned for the different task.

Submit

You got all the answers!

End-of-chapter quiz

1. The emotion dataset contains Twitter messages labeled with emotions. Search for it in the Hub, and read the dataset card. Which of these is not one of its basic emotions?

2. Search for the `ar_sarcasm` dataset in the Hub. Which task does it support?

3. How does the BERT model expect a pair of sentences to be processed?

4. What are the benefits of the `Dataset.map()` method?

5. What does dynamic padding mean?

6. What is the purpose of a collate function?

7. What happens when you instantiate one of the `AutoModelForXxx` classes with a pretrained language model (such as `bert-base-uncased`) that corresponds to a different task than the one for which it was trained?

8. What's the purpose of `TrainingArguments`?

9. Why should you use the `Accelerate` library?

4. What happens when you instantiate one of the `TFAutoModelForXxx` classes with a pretrained language model (such as `bert-base-uncased`) that corresponds to a different task than the one for which it was trained?

5. The TensorFlow models from

LLM Course

Search documentation

EN

3,022

0. SETUP

1. TRANSFORMER MODELS

2. USING TRANSFORMERS

3. FINE-TUNING A PRETRAINED MODEL

Introduction

Processing the data

Fine-tuning a model with the Trainer API or Keras

A full training

Fine-tuning, Check!

End-of-chapter quiz

4. SHARING MODELS AND TOKENIZERS

5. THE DATASETS LIBRARY

6. THE TOKENIZERS LIBRARY

7. CLASSICAL NLP TASKS

8. HOW TO ASK FOR HELP

9. BUILDING AND SHARING DEMOS

10. CURATE HIGH-QUALITY DATASETS

11. FINE-TUNE LARGE LANGUAGE MODELS

12. BUILD REASONING MODELS

COURSE EVENTS

7. What happens when you instantiate one of the `AutoModelForXxx` classes with a pretrained language model (such as `bert-base-uncased`) that corresponds to a different task than the one for which it was trained?

☐ Nothing, but you get a warning.

☒ The head of the pretrained model is discarded and a new head suitable for the task is inserted instead.

Correct! Correct. For example, when we used `AutoModelForSequenceClassification` with `bert-base-uncased`, we got warnings when instantiating the model. The pretrained head is not used for the sequence classification task, so it's discarded and a new head is instantiated with random weights.

☐ The head of the pretrained model is discarded.

☐ Nothing, since the model can still be fine-tuned for the different task.

Submit

You got all the answers!

8. What's the purpose of `TrainingArguments`?

☒ It contains all the hyperparameters used for training and evaluation with the `Trainer`.

Correct! Correct!

☐ It specifies the size of the model.

☐ It just contains the hyperparameters used for evaluation.

☐ It just contains the hyperparameters used for training.

Submit

You got all the answers!

9. Why should you use the `Accelerate` library?

☐ It provides access to faster models.

☐ It provides a high-level API so I don't have to implement my own training loop.

☒ It makes our training loops work on distributed strategies.

Correct! Correct! With `Accelerate`, your training loops will work for multiple GPUs and TPUs.

☐ It provides more optimization functions.

Submit

You got all the answers!

End-of-chapter quiz

1. The emotion dataset contains Twitter messages labeled with emotions. Search for it in the Hub, and read the dataset card. Which of these is not one of its basic emotions?

2. Search for the `ar_sarcasm` dataset in the Hub. Which task does it support?

3. How does the BERT model expect a pair of sentences to be processed?

4. What are the benefits of the `Dataset.map()` method?

5. What does dynamic padding mean?

6. What is the purpose of a collate function?

7. What happens when you instantiate one of the `AutoModelForXxx` classes with a pretrained language model (such as `bert-base-uncased`) that corresponds to a different task than the one for which it was trained?

8. What's the purpose of `TrainingArguments`?

9. Why should you use the `Accelerate` library?

4. What happens when you instantiate one of the `TFAutoModelForXxx` classes with a pretrained language model (such as `bert-base-uncased`) that corresponds to a different task than the one for which it was trained?

5. The TensorFlow models from