Transformers

English | **简体中文** | **繁體中文** | 한국어

Jax, Pytorch, TensorFlow를 위한 최첨단 자연어처리

🤗 Transformers는 분류, 정보 추출, 질문 답변, 요약, 번역, 문장 생성 등을 100개 이상의 언어로 수행할 수 있는 수천개의 사전학습된 모델을 제공합니다. 우리의 목표는 모두가 최첨단의 NLP 기술을 쉽게 사용하는 것입니다.

🤗 Transformers는 이러한 사전학습 모델을 빠르게 다운로드해 특정 텍스트에 사용하고, 원하는 데이터로 fine-tuning해 커뮤니티나 우리의 모델 허브에 공유할 수 있도록 API를 제공합니다. 또한, 모델 구조를 정의하는 각 파이썬 모듈은 완전히 독립적이여서 연구 실험을 위해 손쉽게 수정할 수 있습니다.

🤗 Transformers는 가장 유명한 3개의 딥러닝 라이브러리를 지원합니다. 이들은 서로 완벽히 연동됩니다 — Jax, PyTorch, TensorFlow. 간단하게 이 라이브러리 중 하나로 모델을 학습하고, 또 다른 라이브러리로 추론을 위해 모델을 불러올 수 있습니다.

온라인 데모

대부분의 모델을 모델 허브 페이지에서 바로 테스트해볼 수 있습니다. 공개 및 비공개 모델을 위한 비공개 모델 호스팅, 버전 관리, 추론 API도 제공합니다.

예시:

- BERT로 마스킹된 단어 완성하기
- Electra를 이용한 개체명 인식
- GPT-2로 텍스트 생성하기
- RoBERTa로 자연어 추론하기
- BART를 이용한 요약
- DistilBERT를 이용한 질문 답변
- T5로 번역하기

Transformer와 글쓰기 는 이 저장소의 텍스트 생성 능력에 관한 Hugging Face 팀의 공식 데모입니다.

Hugging Face 팀의 커스텀 지원을 원한다면

퀵 투어

원하는 텍스트에 바로 모델을 사용할 수 있도록, 우리는 `pipeline` API를 제공합니다. Pipeline은 사전학습 모델과 그 모델을 학습할 때 적용한 전처리 방식을 하나로 합칩니다. 다음은 긍정적인 텍스트와 부정적인 텍스트를 분류하기 위해 pipeline을 사용한 간단한 예시입니다:

```
>>> from transformers import pipeline

# Allocate a pipeline for sentiment-analysis
>>> classifier = pipeline('sentiment-analysis')
>>> classifier('We are very happy to introduce pipeline to the transformers repository.')
[{'label': 'POSITIVE', 'score': 0.9996980428695679}]
```

코드의 두번째 줄은 pipeline이 사용하는 사전학습 모델을 다운로드하고 캐시로 저장합니다. 세번째 줄에선 그 모델이 주어진 텍스트를 평가합니다. 여기서 모델은 99.97%의 확률로 텍스트가 긍정적이라고 평가했습니다.

많은 NLP 과제들을 `pipeline`으로 바로 수행할 수 있습니다. 예를 들어, 질문과 문맥이 주어지면 손쉽게 답변을 추출할 수 있습니다:

```
>>> from transformers import pipeline

# Allocate a pipeline for question-answering
>>> question_answerer = pipeline('question-answering')
>>> question_answerer({
...     'question': 'What is the name of the repository ?',
...     'context': 'Pipeline has been included in the huggingface/transformers
repository'
... })
{'score': 0.30970096588134766, 'start': 34, 'end': 58, 'answer':
'huggingface/transformers'}
```

답변뿐만 아니라, 여기에 사용된 사전학습 모델은 확신도와 토크나이즈된 문장 속 답변의 시작점, 끝점까지 반환합니다. 이 튜토리얼에서 `pipeline` API가 지원하는 다양한 과제를 확인할 수 있습니다.

코드 3줄로 원하는 과제에 맞게 사전학습 모델을 다운로드 받고 사용할 수 있습니다. 다음은 PyTorch 버전입니다:

```
>>> from transformers import AutoTokenizer, AutoModel

>>> tokenizer = AutoTokenizer.from_pretrained("bert-base-uncased")
>>> model = AutoModel.from_pretrained("bert-base-uncased")

>>> inputs = tokenizer("Hello world!", return_tensors="pt")
>>> outputs = model(**inputs)
```

다음은 TensorFlow 버전입니다:

```
>>> from transformers import AutoTokenizer, TFAutoModel

>>> tokenizer = AutoTokenizer.from_pretrained("bert-base-uncased")
>>> model = TFAutoModel.from_pretrained("bert-base-uncased")

>>> inputs = tokenizer("Hello world!", return_tensors="tf")
>>> outputs = model(**inputs)
```

토크나이저는 사전학습 모델의 모든 전처리를 책임집니다. 그리고 (위의 예시처럼) 1개의 스트링이나 리스트도 처리할 수 있습니다. 토크나이저는 딕셔너리를 반환하는데, 이는 다운스트림 코드에 사용하거나 언패킹 연산자 ** 를 이용해 모델에 바로 전달할 수도 있습니다.

모델 자체는 일반적으로 사용되는 Pytorch `nn.Module` 나 TensorFlow `tf.keras.Model` 입니다. 이 튜토리얼은 이러한 모델을 표준적인 PyTorch나 TensorFlow 학습 과정에서 사용하는 방법, 또는 새로운 데이터로 fine-tune하기 위해 `Trainer` API를 사용하는 방법을 설명해줍니다.

왜 transformers를 사용해야 할까요?

1. 손쉽게 사용할 수 있는 최첨단 모델:
    - NLU와 NLG 과제에서 뛰어난 성능을 보입니다.
    - 교육자 실무자에게 진입 장벽이 낮습니다.
    - 3개의 클래스만 배우면 바로 사용할 수 있습니다.
    - 하나의 API로 모든 사전학습 모델을 사용할 수 있습니다.
2. 더 적은 계산 비용, 더 적은 탄소 발자국:
    - 연구자들은 모델을 계속 다시 학습시키는 대신 학습된 모델을 공유할 수 있습니다.
    - 실무자들은 학습에 필요한 시간과 비용을 절약할 수 있습니다.
    - 수십개의 모델 구조, 2,000개 이상의 사전학습 모델, 100개 이상의 언어로 학습된 모델 등.
3. 모델의 각 생애주기에 적합한 프레임워크:
    - 코드 3줄로 최첨단 모델을 학습하세요.
    - 자유롭게 모델을 TF2.0나 PyTorch 프레임워크로 변환하세요.
    - 학습, 평가, 공개 등 각 단계에 맞는 프레임워크를 원하는대로 선택하세요.
4. 필요한 대로 모델이나 예시를 커스터마이즈하세요:
    - 우리는 저자가 공개한 결과를 재현하기 위해 각 모델 구조의 예시를 제공합니다.
    - 모델 내부 구조는 가능한 일관적으로 공개되어 있습니다.

- 빠른 실험을 위해 모델 파일은 라이브러리와 독립적으로 사용될 수 있습니다.

왜 transformers를 사용하지 말아야 할까요?

- 이 라이브러리는 신경망 블록을 만들기 위한 모듈이 아닙니다. 연구자들이 여러 파일을 살펴보지 않고 바로 각 모델을 사용할 수 있도록, 모델 파일 코드의 추상화 수준을 적정하게 유지했습니다.
- 학습 API는 모든 모델에 적용할 수 있도록 만들어지진 않았지만, 라이브러리가 제공하는 모델들에 적용할 수 있도록 최적화되었습니다. 일반적인 머신 러닝을 위해선, 다른 라이브러리를 사용하세요.
- 가능한 많은 사용 예시를 보여드리고 싶어서, 예시 폴더의 스크립트를 준비했습니다. 이 스크립트들을 수정 없이 특정한 문제에 바로 적용하지 못할 수 있습니다. 필요에 맞게 일부 코드를 수정해야 할 수 있습니다.

설치

pip로 설치하기

이 저장소는 Python 3.6+, Flax 0.3.2+, PyTorch 1.3.1+, TensorFlow 2.3+에서 테스트 되었습니다.

가상 환경에 🤗 Transformers를 설치하세요. Python 가상 환경에 익숙하지 않다면, 사용자 가이드를 확인하세요.

우선, 사용할 Python 버전으로 가상 환경을 만들고 실행하세요.

그 다음, Flax, PyTorch, TensorFlow 중 적어도 하나는 설치해야 합니다. 플랫폼에 맞는 설치 명령어를 확인하기 위해 TensorFlow 설치 페이지, PyTorch 설치 페이지, Flax 설치 페이지를 확인하세요.

이들 중 적어도 하나가 설치되었다면, 🤗 Transformers는 다음과 같이 pip을 이용해 설치할 수 있습니다:

```
pip install transformers
```

예시들을 체험해보고 싶거나, 최최최첨단 코드를 원하거나, 새로운 버전이 나올 때까지 기다릴 수 없다면 라이브러리를 소스에서 바로 설치하셔야 합니다.

conda로 설치하기

Transformers 버전 v4.0.0부터, conda 채널이 생겼습니다: `huggingface`.

🤗 Transformers는 다음과 같이 conda로 설치할 수 있습니다:

```
conda install -c huggingface transformers
```

Flax, PyTorch, TensorFlow 설치 페이지에서 이들을 conda로 설치하는 방법을 확인하세요.

모델 구조

🤗 **Transformers가 제공하는 모든 모델 체크포인트** 는 huggingface.co 모델 허브에 완벽히 연동되어 있습니다. 개인과 기관이 모델 허브에 직접 업로드할 수 있습니다.

현재 사용 가능한 모델 체크포인트의 개수:

🤗 Transformers는 다음 모델들을 제공합니다 (각 모델의 요약은 여기서 확인하세요):

1. **ALBERT** (from Google Research and the Toyota Technological Institute at Chicago) released with the paper ALBERT: A Lite BERT for Self-supervised Learning of Language Representations, by Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, Radu Soricut.
2. **BART** (from Facebook) released with the paper BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension by Mike Lewis, Yinhan Liu, Naman Goyal, Marjan

Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov and Luke Zettlemoyer.
3. **BARThez** (from École polytechnique) released with the paper BARThez: a Skilled Pretrained French Sequence-to-Sequence Model by Moussa Kamal Eddine, Antoine J.-P. Tixier, Michalis Vazirgiannis.
4. **BARTpho** (from VinAI Research) released with the paper BARTpho: Pre-trained Sequence-to-Sequence Models for Vietnamese by Nguyen Luong Tran, Duong Minh Le and Dat Quoc Nguyen.
5. **BEiT** (from Microsoft) released with the paper BEiT: BERT Pre-Training of Image Transformers by Hangbo Bao, Li Dong, Furu Wei.
6. **BERT** (from Google) released with the paper BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding by Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova.
7. **BERT For Sequence Generation** (from Google) released with the paper Leveraging Pre-trained Checkpoints for Sequence Generation Tasks by Sascha Rothe, Shashi Narayan, Aliaksei Severyn.
8. **BERTweet** (from VinAI Research) released with the paper BERTweet: A pre-trained language model for English Tweets by Dat Quoc Nguyen, Thanh Vu and Anh Tuan Nguyen.
9. **BigBird-Pegasus** (from Google Research) released with the paper Big Bird: Transformers for Longer Sequences by Manzil Zaheer, Guru Guruganesh, Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, Amr Ahmed.
10. **BigBird-RoBERTa** (from Google Research) released with the paper Big Bird: Transformers for Longer Sequences by Manzil Zaheer, Guru Guruganesh, Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, Amr Ahmed.
11. **Blenderbot** (from Facebook) released with the paper Recipes for building an open-domain chatbot by Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M. Smith, Y-Lan Boureau, Jason Weston.
12. **BlenderbotSmall** (from Facebook) released with the paper Recipes for building an open-domain chatbot by Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M. Smith, Y-Lan Boureau, Jason Weston.
13. **BORT** (from Alexa) released with the paper Optimal Subarchitecture Extraction For BERT by Adrian de Wynter and Daniel J. Perry.
14. **ByT5** (from Google Research) released with the paper ByT5: Towards a token-free future with pre-trained byte-to-byte models by Linting Xue, Aditya Barua, Noah Constant, Rami Al-Rfou, Sharan Narang, Mihir Kale, Adam Roberts, Colin Raffel.
15. **CamemBERT** (from Inria/Facebook/Sorbonne) released with the paper CamemBERT: a Tasty French Language Model by Louis Martin*, Benjamin Muller*, Pedro Javier Ortiz Suárez*, Yoann Dupont, Laurent Romary, Éric Villemonte de la Clergerie, Djamé Seddah and Benoît Sagot.
16. **CANINE** (from Google Research) released with the paper CANINE: Pre-training an Efficient Tokenization-Free Encoder for Language Representation by Jonathan H. Clark, Dan Garrette, Iulia Turc, John Wieting.
17. **CLIP** (from OpenAI) released with the paper Learning Transferable Visual Models From Natural Language Supervision by Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever.
18. **ConvBERT** (from YituTech) released with the paper ConvBERT: Improving BERT with Span-based Dynamic Convolution by Zihang Jiang, Weihao Yu, Daquan Zhou, Yunpeng Chen, Jiashi Feng, Shuicheng Yan.
19. **ConvNeXT** (from Facebook AI) released with the paper A ConvNet for the 2020s by Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining Xie.
20. **CPM** (from Tsinghua University) released with the paper CPM: A Large-scale Generative Chinese Pre-trained Language Model by Zhengyan Zhang, Xu Han, Hao Zhou, Pei Ke, Yuxian Gu, Deming Ye, Yujia Qin, Yusheng Su, Haozhe Ji, Jian Guan, Fanchao Qi, Xiaozhi Wang, Yanan Zheng, Guoyang Zeng, Huanqi Cao, Shengqi Chen, Daixuan Li, Zhenbo Sun, Zhiyuan Liu, Minlie Huang, Wentao Han, Jie Tang, Juanzi Li, Xiaoyan Zhu, Maosong Sun.
21. **CTRL** (from Salesforce) released with the paper CTRL: A Conditional Transformer Language Model for Controllable Generation by Nitish Shirish Keskar*, Bryan McCann*, Lav R. Varshney, Caiming Xiong and Richard Socher.
22. **Data2Vec** (from Facebook) released with the paper Data2Vec: A General Framework for Self-supervised Learning in Speech, Vision and Language by Alexei Baevski, Wei-Ning Hsu, Qiantong Xu, Arun Babu, Jiatao Gu, Michael Auli.
23. **DeBERTa** (from Microsoft) released with the paper DeBERTa: Decoding-enhanced BERT with Disentangled Attention by Pengcheng He, Xiaodong Liu, Jianfeng Gao, Weizhu Chen.

24. **DeBERTa-v2** (from Microsoft) released with the paper DeBERTa: Decoding-enhanced BERT with Disentangled Attention by Pengcheng He, Xiaodong Liu, Jianfeng Gao, Weizhu Chen.
25. **Decision Transformer** (from Berkeley/Facebook/Google) released with the paper Decision Transformer: Reinforcement Learning via Sequence Modeling by Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Michael Laskin, Pieter Abbeel, Aravind Srinivas, Igor Mordatch.
26. **DeiT** (from Facebook) released with the paper Training data-efficient image transformers & distillation through attention by Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, Hervé Jégou.
27. **DETR** (from Facebook) released with the paper End-to-End Object Detection with Transformers by Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko.
28. **DialoGPT** (from Microsoft Research) released with the paper DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation by Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, Bill Dolan.
29. **DistilBERT** (from HuggingFace), released together with the paper DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter by Victor Sanh, Lysandre Debut and Thomas Wolf. The same method has been applied to compress GPT2 into DistilGPT2, RoBERTa into DistilRoBERTa, Multilingual BERT into DistilmBERT and a German version of DistilBERT.
30. **DiT** (from Microsoft Research) released with the paper DiT: Self-supervised Pre-training for Document Image Transformer by Junlong Li, Yiheng Xu, Tengchao Lv, Lei Cui, Cha Zhang, Furu Wei.
31. **DPR** (from Facebook) released with the paper Dense Passage Retrieval for Open-Domain Question Answering by Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih.
32. **DPT** (from Intel Labs) released with the paper Vision Transformers for Dense Prediction by René Ranftl, Alexey Bochkovskiy, Vladlen Koltun.
33. **ELECTRA** (from Google Research/Stanford University) released with the paper ELECTRA: Pre-training text encoders as discriminators rather than generators by Kevin Clark, Minh-Thang Luong, Quoc V. Le, Christopher D. Manning.
34. **EncoderDecoder** (from Google Research) released with the paper Leveraging Pre-trained Checkpoints for Sequence Generation Tasks by Sascha Rothe, Shashi Narayan, Aliaksei Severyn.
35. **FlauBERT** (from CNRS) released with the paper FlauBERT: Unsupervised Language Model Pre-training for French by Hang Le, Loïc Vial, Jibril Frej, Vincent Segonne, Maximin Coavoux, Benjamin Lecouteux, Alexandre Allauzen, Benoît Crabbé, Laurent Besacier, Didier Schwab.
36. **FNet** (from Google Research) released with the paper FNet: Mixing Tokens with Fourier Transforms by James Lee-Thorp, Joshua Ainslie, Ilya Eckstein, Santiago Ontanon.
37. **Funnel Transformer** (from CMU/Google Brain) released with the paper Funnel-Transformer: Filtering out Sequential Redundancy for Efficient Language Processing by Zihang Dai, Guokun Lai, Yiming Yang, Quoc V. Le.
38. **GLPN** (from KAIST) released with the paper Global-Local Path Networks for Monocular Depth Estimation with Vertical CutDepth by Doyeon Kim, Woonghyun Ga, Pyungwhan Ahn, Donggyu Joo, Sehwan Chun, Junmo Kim.
39. **GPT** (from OpenAI) released with the paper Improving Language Understanding by Generative Pre-Training by Alec Radford, Karthik Narasimhan, Tim Salimans and Ilya Sutskever.
40. **GPT Neo** (from EleutherAI) released in the repository EleutherAI/gpt-neo by Sid Black, Stella Biderman, Leo Gao, Phil Wang and Connor Leahy.
41. **GPT-2** (from OpenAI) released with the paper Language Models are Unsupervised Multitask Learners by Alec Radford*, Jeffrey Wu*, Rewon Child, David Luan, Dario Amodei** and Ilya Sutskever**.
42. **GPT-J** (from EleutherAI) released in the repository kingoflolz/mesh-transformer-jax by Ben Wang and Aran Komatsuzaki.
43. **Hubert** (from Facebook) released with the paper HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units by Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, Abdelrahman Mohamed.
44. **I-BERT** (from Berkeley) released with the paper I-BERT: Integer-only BERT Quantization by Sehoon Kim, Amir Gholami, Zhewei Yao, Michael W. Mahoney, Kurt Keutzer.
45. **ImageGPT** (from OpenAI) released with the paper Generative Pretraining from Pixels by Mark Chen, Alec Radford, Rewon Child, Jeffrey Wu, Heewoo Jun, David Luan, Ilya Sutskever.
46. **LayoutLM** (from Microsoft Research Asia) released with the paper LayoutLM: Pre-training of Text and Layout for Document Image Understanding by Yiheng Xu, Minghao Li, Lei Cui, Shaohan Huang, Furu Wei, Ming Zhou.

47. **LayoutLMv2** (from Microsoft Research Asia) released with the paper LayoutLMv2: Multi-modal Pre-training for Visually-Rich Document Understanding by Yang Xu, Yiheng Xu, Tengchao Lv, Lei Cui, Furu Wei, Guoxin Wang, Yijuan Lu, Dinei Florencio, Cha Zhang, Wanxiang Che, Min Zhang, Lidong Zhou.
48. **LayoutXLM** (from Microsoft Research Asia) released with the paper LayoutXLM: Multimodal Pre-training for Multilingual Visually-rich Document Understanding by Yiheng Xu, Tengchao Lv, Lei Cui, Guoxin Wang, Yijuan Lu, Dinei Florencio, Cha Zhang, Furu Wei.
49. **LED** (from AllenAI) released with the paper Longformer: The Long-Document Transformer by Iz Beltagy, Matthew E. Peters, Arman Cohan.
50. **Longformer** (from AllenAI) released with the paper Longformer: The Long-Document Transformer by Iz Beltagy, Matthew E. Peters, Arman Cohan.
51. **LUKE** (from Studio Ousia) released with the paper LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention by Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, Yuji Matsumoto.
52. **LXMERT** (from UNC Chapel Hill) released with the paper LXMERT: Learning Cross-Modality Encoder Representations from Transformers for Open-Domain Question Answering by Hao Tan and Mohit Bansal.
53. **M2M100** (from Facebook) released with the paper Beyond English-Centric Multilingual Machine Translation by Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, Armand Joulin.
54. **MarianMT** Machine translation models trained using OPUS data by Jörg Tiedemann. The Marian Framework is being developed by the Microsoft Translator Team.
55. **MaskFormer** (from Meta and UIUC) released with the paper Per-Pixel Classification is Not All You Need for Semantic Segmentation by Bowen Cheng, Alexander G. Schwing, Alexander Kirillov.
56. **MBart** (from Facebook) released with the paper Multilingual Denoising Pre-training for Neural Machine Translation by Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, Luke Zettlemoyer.
57. **MBart-50** (from Facebook) released with the paper Multilingual Translation with Extensible Multilingual Pretraining and Finetuning by Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, Angela Fan.
58. **Megatron-BERT** (from NVIDIA) released with the paper Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism by Mohammad Shoeybi, Mostofa Patwary, Raul Puri, Patrick LeGresley, Jared Casper and Bryan Catanzaro.
59. **Megatron-GPT2** (from NVIDIA) released with the paper Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism by Mohammad Shoeybi, Mostofa Patwary, Raul Puri, Patrick LeGresley, Jared Casper and Bryan Catanzaro.
60. **mLUKE** (from Studio Ousia) released with the paper mLUKE: The Power of Entity Representations in Multilingual Pretrained Language Models by Ryokan Ri, Ikuya Yamada, and Yoshimasa Tsuruoka.
61. **MPNet** (from Microsoft Research) released with the paper MPNet: Masked and Permuted Pre-training for Language Understanding by Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, Tie-Yan Liu.
62. **MT5** (from Google AI) released with the paper mT5: A massively multilingual pre-trained text-to-text transformer by Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, Colin Raffel.
63. **Nyströmformer** (from the University of Wisconsin - Madison) released with the paper Nyströmformer: A Nyström-Based Algorithm for Approximating Self-Attention by Yunyang Xiong, Zhanpeng Zeng, Rudrasis Chakraborty, Mingxing Tan, Glenn Fung, Yin Li, Vikas Singh.
64. **Pegasus** (from Google) released with the paper PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization by Jingqing Zhang, Yao Zhao, Mohammad Saleh and Peter J. Liu.
65. **Perceiver IO** (from Deepmind) released with the paper Perceiver IO: A General Architecture for Structured Inputs & Outputs by Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, Olivier Hénaff, Matthew M. Botvinick, Andrew Zisserman, Oriol Vinyals, João Carreira.
66. **PhoBERT** (from VinAI Research) released with the paper PhoBERT: Pre-trained language models for Vietnamese by Dat Quoc Nguyen and Anh Tuan Nguyen.

67. **PLBart** (from UCLA NLP) released with the paper Unified Pre-training for Program Understanding and Generation by Wasi Uddin Ahmad, Saikat Chakraborty, Baishakhi Ray, Kai-Wei Chang.
68. **PoolFormer** (from Sea AI Labs) released with the paper MetaFormer is Actually What You Need for Vision by Yu, Weihao and Luo, Mi and Zhou, Pan and Si, Chenyang and Zhou, Yichen and Wang, Xinchao and Feng, Jiashi and Yan, Shuicheng.
69. **ProphetNet** (from Microsoft Research) released with the paper ProphetNet: Predicting Future N-gram for Sequence-to-Sequence Pre-training by Yu Yan, Weizhen Qi, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang and Ming Zhou.
70. **QDQBert** (from NVIDIA) released with the paper Integer Quantization for Deep Learning Inference: Principles and Empirical Evaluation by Hao Wu, Patrick Judd, Xiaojie Zhang, Mikhail Isaev and Paulius Micikevicius.
71. **REALM** (from Google Research) released with the paper REALM: Retrieval-Augmented Language Model Pre-Training by Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat and Ming-Wei Chang.
72. **Reformer** (from Google Research) released with the paper Reformer: The Efficient Transformer by Nikita Kitaev, Łukasz Kaiser, Anselm Levskaya.
73. **RegNet** (from META Research) released with the paper Designing Network Design Space by Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, Piotr Dollár.
74. **RemBERT** (from Google Research) released with the paper Rethinking embedding coupling in pre-trained language models by Hyung Won Chung, Thibault Févry, Henry Tsai, M. Johnson, Sebastian Ruder.
75. **ResNet** (from Microsoft Research) released with the paper Deep Residual Learning for Image Recognition by Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun.
76. **RoBERTa** (from Facebook), released together with the paper a Robustly Optimized BERT Pretraining Approach by Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov.
77. **RoFormer** (from ZhuiyiTechnology), released together with the paper a RoFormer: Enhanced Transformer with Rotary Position Embedding by Jianlin Su and Yu Lu and Shengfeng Pan and Bo Wen and Yunfeng Liu.
78. **SegFormer** (from NVIDIA) released with the paper SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers by Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, Ping Luo.
79. **SEW** (from ASAPP) released with the paper Performance-Efficiency Trade-offs in Unsupervised Pre-training for Speech Recognition by Felix Wu, Kwangyoun Kim, Jing Pan, Kyu Han, Kilian Q. Weinberger, Yoav Artzi.
80. **SEW-D** (from ASAPP) released with the paper Performance-Efficiency Trade-offs in Unsupervised Pre-training for Speech Recognition by Felix Wu, Kwangyoun Kim, Jing Pan, Kyu Han, Kilian Q. Weinberger, Yoav Artzi.
81. **SpeechToTextTransformer** (from Facebook), released together with the paper fairseq S2T: Fast Speech-to-Text Modeling with fairseq by Changhan Wang, Yun Tang, Xutai Ma, Anne Wu, Dmytro Okhonko, Juan Pino.
82. **SpeechToTextTransformer2** (from Facebook), released together with the paper Large-Scale Self- and Semi-Supervised Learning for Speech Translation by Changhan Wang, Anne Wu, Juan Pino, Alexei Baevski, Michael Auli, Alexis Conneau.
83. **Splinter** (from Tel Aviv University), released together with the paper Few-Shot Question Answering by Pretraining Span Selection by Ori Ram, Yuval Kirstain, Jonathan Berant, Amir Globerson, Omer Levy.
84. **SqueezeBert** (from Berkeley) released with the paper SqueezeBERT: What can computer vision teach NLP about efficient neural networks? by Forrest N. Iandola, Albert E. Shaw, Ravi Krishna, and Kurt W. Keutzer.
85. **Swin Transformer** (from Microsoft) released with the paper Swin Transformer: Hierarchical Vision Transformer using Shifted Windows by Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo.
86. **T5** (from Google AI) released with the paper Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer by Colin Raffel and Noam Shazeer and Adam Roberts and Katherine Lee and Sharan Narang and Michael Matena and Yanqi Zhou and Wei Li and Peter J. Liu.
87. **T5v1.1** (from Google AI) released in the repository google-research/text-to-text-transfer-transformer by Colin Raffel and Noam Shazeer and Adam Roberts and Katherine Lee and Sharan Narang and Michael Matena and Yanqi Zhou and Wei Li and Peter J. Liu.
88. **TAPAS** (from Google AI) released with the paper TAPAS: Weakly Supervised Table Parsing via Pre-training by Jonathan Herzig, Paweł Krzysztof Nowak, Thomas Müller, Francesco Piccinno and Julian Martin Eisenschlos.

89. **Transformer-XL** (from Google/CMU) released with the paper Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context by Zihang Dai*, Zhilin Yang*, Yiming Yang, Jaime Carbonell, Quoc V. Le, Ruslan Salakhutdinov.
90. **TrOCR** (from Microsoft), released together with the paper TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models by Minghao Li, Tengchao Lv, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, Furu Wei.
91. **UniSpeech** (from Microsoft Research) released with the paper UniSpeech: Unified Speech Representation Learning with Labeled and Unlabeled Data by Chengyi Wang, Yu Wu, Yao Qian, Kenichi Kumatani, Shujie Liu, Furu Wei, Michael Zeng, Xuedong Huang.
92. **UniSpeechSat** (from Microsoft Research) released with the paper UNISPEECH-SAT: UNIVERSAL SPEECH REPRESENTATION LEARNING WITH SPEAKER AWARE PRE-TRAINING by Sanyuan Chen, Yu Wu, Chengyi Wang, Zhengyang Chen, Zhuo Chen, Shujie Liu, Jian Wu, Yao Qian, Furu Wei, Jinyu Li, Xiangzhan Yu.
93. **VAN** (from Tsinghua University and Nankai University) released with the paper Visual Attention Network by Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, Shi-Min Hu.
94. **ViLT** (from NAVER AI Lab/Kakao Enterprise/Kakao Brain) released with the paper ViLT: Vision-and-Language Transformer Without Convolution or Region Supervision by Wonjae Kim, Bokyung Son, Ildoo Kim.
95. **Vision Transformer (ViT)** (from Google AI) released with the paper An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale by Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby.
96. **VisualBERT** (from UCLA NLP) released with the paper VisualBERT: A Simple and Performant Baseline for Vision and Language by Liunian Harold Li, Mark Yatskar, Da Yin, Cho-Jui Hsieh, Kai-Wei Chang.
97. **ViTMAE** (from Meta AI) released with the paper Masked Autoencoders Are Scalable Vision Learners by Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, Ross Girshick.
98. **Wav2Vec2** (from Facebook AI) released with the paper wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations by Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, Michael Auli.
99. **Wav2Vec2Phoneme** (from Facebook AI) released with the paper Simple and Effective Zero-shot Cross-lingual Phoneme Recognition by Qiantong Xu, Alexei Baevski, Michael Auli.
100. **WavLM** (from Microsoft Research) released with the paper WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing by Sanyuan Chen, Chengyi Wang, Zhengyang Chen, Yu Wu, Shujie Liu, Zhuo Chen, Jinyu Li, Naoyuki Kanda, Takuya Yoshioka, Xiong Xiao, Jian Wu, Long Zhou, Shuo Ren, Yanmin Qian, Yao Qian, Jian Wu, Michael Zeng, Furu Wei.
101. **XGLM** (From Facebook AI) released with the paper Few-shot Learning with Multilingual Language Models by Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, Mona Diab, Veselin Stoyanov, Xian Li.
102. **XLM** (from Facebook) released together with the paper Cross-lingual Language Model Pretraining by Guillaume Lample and Alexis Conneau.
103. **XLM-ProphetNet** (from Microsoft Research) released with the paper ProphetNet: Predicting Future N-gram for Sequence-to-Sequence Pre-training by Yu Yan, Weizhen Qi, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang and Ming Zhou.
104. **XLM-RoBERTa** (from Facebook AI), released together with the paper Unsupervised Cross-lingual Representation Learning at Scale by Alexis Conneau*, Kartikay Khandelwal*, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer and Veselin Stoyanov.
105. **XLM-RoBERTa-XL** (from Facebook AI) released with the paper Larger-Scale Transformers for Multilingual Masked Language Modeling by Naman Goyal, Jingfei Du, Myle Ott, Giri Anantharaman, Alexis Conneau.
106. **XLNet** (from Google/CMU) released with the paper XLNet: Generalized Autoregressive Pretraining for Language Understanding by Zhilin Yang*, Zihang Dai*, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, Quoc V. Le.
107. **XLS-R** (from Facebook AI) released with the paper XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale by Arun Babu, Changhan Wang, Andros Tjandra, Kushal Lakhotia, Qiantong Xu, Naman Goyal, Kritika Singh, Patrick von Platen, Yatharth Saraf, Juan Pino, Alexei Baevski, Alexis Conneau, Michael Auli.
108. **XLSR-Wav2Vec2** (from Facebook AI) released with the paper Unsupervised Cross-Lingual Representation Learning For Speech Recognition by Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, Michael Auli.

109. **YOSO** (from the University of Wisconsin - Madison) released with the paper [You Only Sample (Almost) by Zhanpeng Zeng, Yunyang Xiong, Sathya N. Ravi, Shailesh Acharya, Glenn Fung, Vikas Singh.
110. 새로운 모델을 올리고 싶나요? 우리가 **상세한 가이드와 템플릿** 으로 새로운 모델을 올리도록 도와드릴게요. 가이드와 템플릿은 이 저장소의 `templates` 폴더에서 확인하실 수 있습니다. 컨트리뷰션 가이드라인을 꼭 확인해주시고, PR을 올리기 전에 메인테이너에게 연락하거나 이슈를 오픈해 피드백을 받으시길 바랍니다.

각 모델이 Flax, PyTorch, TensorFlow으로 구현되었는지 또는 🤗 Tokenizers 라이브러리가 지원하는 토크나이저를 사용하는지 확인하려면, 이 표를 확인하세요.

이 구현은 여러 데이터로 검증되었고 (예시 스크립트를 참고하세요) 오리지널 구현의 성능과 같아야 합니다. 도큐먼트의 Examples 섹션에서 성능에 대한 자세한 설명을 확인할 수 있습니다.

더 알아보기

섹션	설명
도큐먼트	전체 API 도큐먼트와 튜토리얼
과제 요약	🤗 Transformers가 지원하는 과제들
전처리 튜토리얼	`Tokenizer` 클래스를 이용해 모델을 위한 데이터 준비하기
학습과 fine-tuning	🤗 Transformers가 제공하는 모델 PyTorch/TensorFlow 학습 과정과 `Trainer` API에서 사용하기
퀵 투어: Fine-tuning/사용 스크립트	다양한 과제에서 모델 fine-tuning하는 예시 스크립트
모델 공유 및 업로드	커뮤니티에 fine-tune된 모델을 업로드 및 공유하기
마이그레이션	`pytorch-transformers`나 `pytorch-pretrained-bert`에서 🤗 Transformers로 이동하기

인용

🤗 Transformers 라이브러리를 인용하고 싶다면, 이 논문을 인용해 주세요:

```
@inproceedings{wolf-etal-2020-transformers,
    title = "Transformers: State-of-the-Art Natural Language Processing",
    author = "Thomas Wolf and Lysandre Debut and Victor Sanh and Julien Chaumond and Clement Delangue and Anthony Moi and Pierric Cistac and Tim Rault and Rémi Louf and Morgan Funtowicz and Joe Davison and Sam Shleifer and Patrick von Platen and Clara Ma and Yacine Jernite and Julien Plu and Canwen Xu and Teven Le Scao and Sylvain Gugger and Mariama Drame and Quentin Lhoest and Alexander M. Rush",
    booktitle = "Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations",
    month = oct,
    year = "2020",
    address = "Online",
    publisher = "Association for Computational Linguistics",
    url = "https://www.aclweb.org/anthology/2020.emnlp-demos.6",
    pages = "38--45"
}
```