|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 模型名称 | 模型结构 | 方向 | 编码 | 预训练任务 | 模型应用 | 模型结构2 | Tok/  Span/  Sent | 训练语料 | 模型大小 |  |
| BERT | Bert | Mask |  | character-level BPE | MLM/NSP | NLU | Transformer Enc | Tok | Wijipedia  Bookcorpus | ≈500M | http://arxiv.org/abs/1810.04805 |
| SpanBERT | Mask | Bidi | character-level BPE | MLM/SBO | NLU | Transformer Enc | Span | Wijipedia  Bookcorpus | ≈500M | https://arxiv.org/pdf/1907.10529.pdf |
| RoBERTa | 动态Mask |  | byte BPE | MLM | NLU | Transformer Enc | Tok | BookCorpus,  En-Wikipedia  CC-News  OpenWebText  Stories | ≈500M | http://arxiv.org/abs/1907.11692 |
| StructBERT |  |  |  | MLM/word strcture (单词纠正)/sentence structure （句子顺序） |  |  |  |  | ≈1G （StructBERT-large） | https://arxiv.org/abs/1908.04577 |
| ALBERT | Mask |  |  | MLM/SOP |  | 所有层参数共享 |  | Wijipedia  Bookcorpus | ≈100M | http://arxiv.org/abs/1909.11942 |
| KnowBERT |  |  |  |  |  |  |  |  |  |  |
| DeBERTa | Mask |  |  | MLM/解耦注意力/虚拟对抗训练 | NLU |  | Tok |  |  | 2111.09543 |
| K-BERT |  |  |  |  |  |  |  |  |  | arxiv.org/abs/1909.07606v1 |
| K-Adapter |  |  |  |  |  |  |  |  |  |  |
| TaBERT |  |  |  |  |  |  |  |  |  |  |
| 特定领域 | BIOBERT |  |  |  |  |  |  |  | Wijipedia  Bookcorpus  PubMed  PMC | ≈400M | 2019 |
| SCIBERT |  |  |  |  |  |  |  | Semantic Scholar ( full text of the papers in training) | ≈400M | https://www.aclweb.org/anthology/D19-1371 |
| ClinicalBERT |  |  |  |  |  |  |  |  |  |  |
| Bio\_ClinicalBERT |  |  | WordPiece |  |  |  |  | MIMIC III | ≈1G | https://arxiv.org/abs/1903.10676 |
| FinBERT | Mask |  |  | CTR/CWP/SDP/TPP | Fin-NLU |  | Span |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| 特定任务 | SentiLR |  |  |  |  |  |  |  |  |  |  |
| GSG |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| 知识蒸馏 | DistilBERT |  |  |  |  |  |  |  |  |  |  |
| TinyBERT |  |  |  |  |  |  |  |  |  |  |
| MobileBERT |  |  |  |  |  |  |  |  |  |  |
|  | XLNet |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | ERNIE-T |  |  |  |  |  |  |  |  |  |  |
|  | ERNIE-B |  |  |  |  |  |  |  |  |  |  |
|  | ERNIE-NG |  |  |  |  |  |  |  |  |  |  |
|  | 模型名称 | 模型结构 | 方向 | 编码 | 预训练任务 | 模型应用 | 模型结构2 | Tok/  Span/  Sent | 训练语料 | 模型大小 |  |
|  | ERNIE-B2 |  |  |  |  |  |  |  |  |  |  |
|  | ERNIE-M |  |  |  |  |  |  |  |  |  |  |
|  | ERNIE-B3 |  |  |  |  |  |  |  |  |  |  |
|  | GPT-1 |  |  |  |  |  |  |  |  |  |  |
|  | GPT-2 |  |  |  |  |  |  |  |  |  |  |
|  | GPT-3 |  |  |  |  |  |  |  |  |  |  |
|  | MASS | Mask |  |  | CTR | NLG | Transformer-Enc-Dec | Span |  |  | https://arxiv.org/pdf/1905.02450.pdf |
|  | BART | Mask | Bidirectional |  | CTR (Token masking/Token deletion/ Text infilling/ Sentence Permutation/ Document Rotation | NLG | Transformer-Enc-Dec |  |  | ≈500M | http://arxiv.org/abs/1910.13461 |
|  | 模型名称 | 模型结构 | 方向 | 编码 | 预训练任务 | 模型应用 | 模型结构2 | Tok/  Span/  Sent | 训练语料 | 模型大小 |  |
|  | T5 |  |  |  |  | NLG |  |  | C4  Wiki-DPR | ≈1G |  |
|  | ByT5 |  |  |  |  |  |  |  | mC4 |  | https://arxiv.org/abs/2105.13626 |
| 多语言模型 | XNLG |  |  |  | MLM (Mask Langage Model)  DAE (Denoising Auto-Encoding)  XMLM (Cross Lingual MLM)  XAE (Cross-Lingual DAE) | NLG | Transformer-Enc-Dec |  |  |  |  |
| XLM |  |  | 字节对编码（BPE） | CLM (Causal Language Model)  MLM (Masked Language Model)  TLM (Translation Language Model) |  |  |  |  |  | 2019 |
| XLM-R |  |  |  |  |  |  |  |  |  |  |
| XNM-E |  |  |  |  |  |  |  |  |  |  |
|  | UniLM |  |  |  | unidirectional prediction  bidirectional prediction  seuqnece-to-sequence prediction | NLG  NLU |  | Span |  |  | 2019 |
|  | CPM ( (Chinese Pre-trained Language Model)) |  |  |  |  |  | Transformer-based autoregressive language model |  | 100G 中文语料 | ≈10G | <https://huggingface.co/models>?  arxiv=arxiv:2012.00413 |
|  | CPM-2 |  |  |  |  |  |  |  |  |  |  |
|  | XLNet | Mask |  |  | (PLM) Permutation Language Model | NLG |  |  | Wikipedia  Bookcorpus | ≈500M | http://arxiv.org/abs/1906.08237 |
| 多模态模型 | M6 |  |  |  |  |  |  |  |  |  |  |

常用预训练模型总结

只统计单语言、单模态PLM

NLG：自然语言生成

NLU：自然语言理解