Sci, Bio, Fin, K分别代表科学、生物医学、金融和知识。Tok, Sent, Doc分别表示token、句子和文档。Region, Frame分别表示图像和视频的基本单位。Standard language model（SLM）的目标正是如此，训练模型以优化训练语料库中文本的概率P（x）（Radford等人，2019）。在这些情况下，文本通常以自回归的方式进行预测，每次预测序列中的token。这通常是从左到右进行的（详见下文），但也可以按其他顺序进行。

损坏的文本重建Corrupted Text Reconstruction（CTR） 这些目标通过仅对输入句子的噪声部分计算损失，将处理后的文本恢复到未损坏的状态。

全文重构 Full Text Reconstruction（FTR） 这些目标通过计算整个输入文本的损失来重构文本，无论其是否经过噪声处理（刘易斯等人，2020a）。

从左到右的LM（L2R LM）是自回归式LM的一种

LM : Mask; L2R (Left to Right);前缀语言模型;编码器和解码器(Encoder-Decoder)

预训练模型辅助目标：

Next Sentence Prediction (NSP) (Devlin et al., 2019)：一种二元分类损失，用于预测两个片段是否连续出现在较大的文档中，或者是随机不相关的句子。

Sentence Order Prediction (SOP)（Lan 等人，2020 年）：用于预测两个句子是自然顺序还是交换顺序的二元分类损失。

Capital Word Prediction (CWP)（Liu 等人，2020b）：对每个词计算的二进制分类目标，预测每个词是否大写。

Sentence Deshuffling (SDS) (Liu et al., 2020b)：一种用于重组置换段的多类分类任务。

Sentence distance prediction句子距离预测 (SDP) (Liu et al., 2020b)：三类分类任务，预测两个句子之间的位置关系（在同一个文档中相邻，在同一个文档中不相邻，在不同文档中）。

Masked Column Prediction (MCP) (屏蔽列预测 (MCP)（Yin 等人，2020 年）：给定一个表，恢复屏蔽列的名称和数据类型。 Linguistic-Visual Alignment (LVA) 语言视觉对齐 (LVA)（Lu 等人，2019 年）：预测文本内容是否可以与视觉内容对齐的二元分类。

Image Region prediction (IRP)图像区域预测 (IRP) (Su et al., 2020)：给定一个部分特征被屏蔽（归零）的图像，预测被屏蔽的区域。

Replaced Token Detection (RTD)替换令牌检测 (RTD) (Xiao et al., 2021)：预测损坏输入中的每个令牌是否被生成样本替换的二进制分类损失。

Discourse Relation Prediction (DRP)话语关系预测 (DRP) (Sun et al., 2020)：预测两个句子之间的语义或修辞关系。

Translation Language Modeling (TLM)翻译语言建模 (TLM)（Lample 和 Conneau，2019 年）：在源句和目标句中随机考虑平行句和掩码词。

Information Retrieval Relevance信息检索相关性 (IRR) (Sun et al., 2020)：预测两个句子的信息检索相关性。

Token-Passage Prediction (TPP)（Liu 等人，2020b）：识别出现在片段中的段落的关键字。

Universal Knowledge-Text Prediction通用知识文本预测 (UKTP)（Sun 等人，2021 年）：将知识整合到一个预训练的语言模型中。

Machine Translation (机器翻译 (MT) (Chi et al., 2021a)：将句子从源语言翻译成目标语言。

Translation Pair Span Corruption翻译对跨度损坏 (TPSC) (Chi et al., 2021a)：预测翻译对的掩码跨度。

Translation Span Corruption翻译跨度损坏 (TSC) (Chi et al., 2021a)：与 TPSC 不同，TSC 仅掩盖和预测一种语言的跨度

Multilingual Replaced Token Detection多语言替换标记检测 (MRTD) (Chi et al., 2021b)：通过生成对抗网络区分真实输入标记和损坏的多语言句子，其中生成器和鉴别器在语言之间共享。

Translation Replaced Token Detection翻译替换标记检测 (TRTD) (Chi et al., 2021b)：通过生成对抗网络区分翻译对中的真实标记和掩码标记。

Knowledge Embedding知识嵌入 (KE)（Wang 等人，2021 年）：将知识图 (KG) 中的实体和关系编码为分布式表示

Image-to-text transfer图像到文本传输 (ITT)（Wang 等人，2021 年）：类似于为输入图像生成相应描述的图像标题。

Multimodality-to-text transfer (MTT) (Wang et al., 2021)：根据视觉信息和噪声语言信息生成目标文本

应用：

NLG——自然语言生成

NLU：自然语言理解 (Natural Language Understanding)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Family | Models |  | LM | Main | Auxiliary(辅助) | Parallel | Mask | Replace | Delete | Permute | Application |
| GPT | GPT[139] |  | L2R | SLM | - | x | - |  |  |  | NLG |
| GPT-2[1401 |  | L2R | SLM | - | x | - |  |  |  | NLG |
| GPT-3[16] |  | L2R | SLM | - | x | - | - | - | - | NLG |
| Codex[20] |  | L2R | SLM | - | x | - | - | - | - | NLG |
| ELMO | ELMo[130] |  | L2R | SLM |  | x | - |  | - | - | NLU.NLG |
| Bert | BERT[32] |  | Mask | CTR | NSP  (Next Sentence Prediction) | x | Tok |  | - | - | NLU |
| RoBERTa[105] |  | Mask | CTR | 、 | x | Tok |  | - | - | NLU |
| SpanBERT[70] |  | Mask | CTR | - | x | Span |  | - | - | NLU |
| DeBERTa[60] |  | Mask | CTR | - | x | Tok |  | - | - | NLU |
| SciBERT[7] |  | Mask | CTR | NSP | x | Tok |  | - |  | Sci-NLU |
| BioBERT[89] |  | Mask | CTR | NSP | x | Tok |  | - |  | Bio-NLU |
| ALBERT[87] |  | Mask | CTR | SOP  (sentence order prediction) | x | Tok |  |  |  | mSent |
| FinBERT[108] |  | Mask | CTR | CWP.SDS.  CWP：首字母大写预测；  SDS：句子洗牌 | X | Span |  | - | Sent | Fin-NLU |
|  | SDP.TPP  SDP：句子距离预测  TPP：识别出现在片段中的段落的关键词 |
| VLBERT[164] |  | Mask | CTR | IRP | ✓ | Tok.Region |  | - |  | VLU |
| ViLBERT[110] |  | Mask | CTR | IRP.LVA  LVA：语言视觉对齐 | ✓ | Tok,Region |  |  |  | VLU |
| BEIT[5] |  | Mask | CTR.FTR | - | x | Visual"Tok" |  | - | - | VLU |
| VideoBERT[166] |  | Mask | CTR | LVA | ✓ | Tok.Frame |  | - | - | VLU |
| TaBERT[189] |  | Mask | CTR | MCP  MCP：掩蔽列预测 | ✓ | Tok.Column |  | - | - | Tab2Text |
| mBERT[32] |  | Mask | CTR | NSP | x | Tok |  | - | - | XLU |
| TinyBERT[69] |  | Mask | CTR | NSP | x | Tok |  | - | - | XLU |
| ERNIE | ERNIE-T[199] |  | Mask | CTR | NSP | x | Tok.Entity |  | - | - | NLU |
| ERNIE-B[169] |  | Mask | CTR | - | x | Tok.Entity,Phrase |  | - | - | NLU |
| ERNIE-NG[183] |  | Mask | CTR | RTD  (replaced token prediction) | x | N-gram | Tok | - | - | NLU |
| ERNIE-B2[168] |  | Mask | CTR | CWP.SDS,SDP. | x | Entity,Phrase |  | - | Sent | NLU |
|  | SDP.DRP.IRR |
| ERNIE-M[126] |  | LPM | CTR | - | ✓ | Tok |  | - |  | XLU.XLG |
| ERNIE-B3[167] |  | Mask | CTR | SOP.SDP.UKTP | x | Entity,Phrase |  | - | - | NLU |
| BART | BART[94] |  | En-De | FTR | - | x | Tok | Span | Tok | Sent.Doc | NLU,NLG |
| mBART[104] |  | En-De | FTR |  | x | Span |  | - | Sent | NLG |
| UniLM | UniLM1[35] |  | LPM | SLM.CTR | NSP | x | Tok | - | - | - | NLU,NLG |
| UniLM2[6] |  | LPM | SLM.CTR | - | x | Tok | - | - | Tok | NLU,NLG |
| T5 | T5[141] |  | En-De | CTR | - | x | - | Span | - | - | NLU,NLG |
| mT5[186] |  | En-De | CTR | - | x | - | Span | - | - | XLU.XLG |
| mT6[22] |  | En-De | CTR | MT.TPSC.TSC | ✓ | - | Span | - | - | XLU.XLG |
| ByT5[185] |  | En-De | CTR | - | x | - | byte-span | - | - | XLU.XLG |
| XLM | XLM[86] |  | LPM | CTR | TLM | ✓ | Tok | - | - | - | XLU.XLG |
| XLM-R[28] |  | Mask | CTR | - | x | Tok | - | 、 | - | XLU |
| XLM-E[23] |  | Mask | CTR | MRTD,TRTD | X | - | Tok | - | - | XLU,XLG |
| CPM | CPM[200] |  | L2R | SLM | - | x | - | - | - | - | NLG |
| CPM-2[198] |  | En-De | CTR | - | x | Span | - | - | - | NLU,NLG |
| Other | XLNet[188] |  | L2R | SLM | - | x |  |  | - | Tok | NLU |
| PanGu-a[194] |  | L2R | SLM |  | x | - |  | - | - | NLG |
| ELECTRA[26] |  | Mask | CTR | RTD | x | Tok | Tok | - |  | NLU.NLG |
| MASS[162] |  | En-De | CTR | - | x | Span |  | - | - | NLG |
| PEGASUS[195] |  | En-De | CTR |  | x | Tok.Sent |  | - | - | Summarization |
| M6[179] |  | En-De | CTR | ITT.MTT | x | Span | - | - | - | NLG |

|  |  |  |
| --- | --- | --- |
| 任务范式 | 具体任务 | 任务解释 |
| 分类任务 | 文本分类（text classification） |  |
| 自然语言推理（Natural language Inference, NLI） | NLI 旨在预测两个给定句子的关系（例如蕴涵）。 |
| 信息抽取 | 关系抽取（relation extraction） |  |
| 语义解析 (semantic parsing) | 语义解析语义解析是在给定自然语言输入的情况下生成结构化含义表示的任务。 |
| 命名实体识别 (Named entity recognition) |  |
| 语言推理 | 常识推理 (Commonsense Reasoning) | (在上下文识别模棱两可的代词的先行词，如一个例子可能是“奖杯不适合棕色手提箱，因为它太大了。”该模型的任务是推断“它”是指奖杯还是“手提箱”。 |
| 在给定多个选择的情况下完成一个句子。一个例子可能是“埃莉诺主动为她的访客准备了一些咖啡。然后她意识到她没有干净的[Z]。”。候选选项是“杯子”、“碗”和“勺子”。预训练 LM 的任务是从三个候选者中选择最符合常识的一个。 |
| 数学推理 (Mathematical Reasoning) | 解决数学问题的能力，比如算数加法、函数评估 |
| 问答任务  (Question Answering) | 抽取式QA | 从包含答案的上下文文档中识别内容 |
| SQuAD |  |
| 多项选择QA |  |
| RACE |  |
| 自由形式QA |  |
| 文本生成任务  (Text Generation) | 翻译任务 |  |
| 文本摘要 |  |
| 其他类文本生成 |  |

表2：代表性PTMs

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| PTMs | Architecture | Input | Pre-Training Task | Corpus | Params | GLUE | FT? |
| ELMo[14] | LSTM | Text | BiLM | WikiText-103 |  |  | No |
| GPT[15] | Transformer Dec. | Text | LM | BookCorpus | 117M | 72.8 | Yes |
| GPT-2[58] | Transformer Dec. | Text | LM | WebText | 117M~1542M |  | No |
| BERT[16] | Transformer Enc. | Text | MLM&NSP | WikiEn+BookCorpus | 110M~340M | 81.9\* | Yes |
| InfoWord[55] | Transformer Enc. | Text | DIM+MLM | WikiEn+BookCorpus | =BERT | 81.1\* | Yes |
| RoBERTa[43] | Transformer Enc. | Text | MLM | BookCorpus+CC- | 355M | 88.5 | Yes |
| XLNet[49] |  |  |  | News+Open WebText+STORIES |  |  |  |
| Two-Stream | Text | PLM | WikiEn+BookCorpus+Giga5 | ≈BERT | 90.5% | Yes |
| Transformer Enc. |  |  | +Clue Web+Common Crawl |  |  |  |
| ELECTRA[56] | Transformer Enc. | Text | RTD+MLM | same to XLNet | 335M | 88.6 | Yes |
| UniLM[44] |  | Text | MLM+ NSP | WikiEn+BookCorpus | 340M | 80.8 | Yes |
| MASS[41] | Transformer | Text | Seq2Seq MLM | \*Task-dependent |  |  | Yes |
| BART[50] | Transformer | Text | DAE | same to RoBERTa | 110% of BERT | 88.4\* | Yes |
| T5[42] | Transformer | Text | Seq2Seq MLM | Colossal Clean Crawled Corpus(C4) | 220M~11B | 89.7\* | Yes |
| ERNIE(THU)[76] | Transformer Enc. | Text+Entities | MLM+NSP+dEA | WikiEn+Wikidata | 114M | 79.6 | Yes |
| KnowBERT[77] | Transformer Enc. | Text | MLM+NSP+EL | WikiEn+WordNet/Wiki | 253M~523M |  | Yes |
| K-BERT[78] | Transformer Enc. | Text+Triples | MLM+NSP | WikiZh+WebtextZh+CN-DBpedia | =BERT |  | Yes |
|  |  |  | HowNet+MedicalKG |  |  |  |
| KEPLER[80] | Transformer Enc. | Text | MLM+KE | WikiEn+Wikidata/WordNet |  |  | Yes |
| WKLM[57] | Transformer Enc. | Text | MLM+ERD | WikiEn+Wikidata | =BERT |  | Yes |
| CoLAKE[81] | Transformer Enc. | Text+Triples | MLM | WikiEn+Wikidata | =RoBERTa | 86.3 | Yes |