

一、在预训练模型中引入知识的原因

- 1. 在如BERT的预训练模型中，存在类似于Masked Language Model的预训练任务，但是如果一个句子中存在人名、地名或者其他信息的时候，mask机制可能破坏完整的实体信息。（e.g Horry Poter只遮挡Horry）
- 2. 引入外部的知识，例如一个句子中出现了太阳，则可以从知识图谱中“联想”到地球，金星等信息。

二、在预训练任务中引入知识的方式

- 1. 知识库固定
- (1) 在预训练任务中引入知识；
- (2) 将entity的embedding线性/非线性变换到word embedding空间中，然后嵌入到文本中；
- (3) 设计基于BERT的辅助结构编码知识；
- 3. 知识库和LM联合训练。

模型名称	时间及会议	基本结构	知识引入方式	创新点	预训练任务	提升领域
ERNIE1.0	ACL2019	Transformer Enc	预训练任务（实体MLM）	改进BERT预训练任务，在预训练任务中引入知识	Basic-Level Masking+ Phrase-Level Masking+ Entity-Level Masking+ Dialogue Language Model	NLU
LIBERT	arXiv2019	Transformer Enc	预训练任务（词汇关系）	将语义相似度知识引入预训练任务	Masked Language Model+ Next Sentence Predict+ Lexical Relation Classifier	NLU
Sense BERT	arXiv2019	Transformer Enc	预训练任务（语义信息）	将外部语义信息引入预训练任务	Masked Language Model(with word sense information)+ Next Sentence Predict+	语义消歧+少部分NLU
SentiLARE	EMNLP2020	Transformer Enc	预训练任务（词性标注）	引入了单词级的语言知识，包括词性标注和先验情感极性	Next Sentence Predict+ label-aware masked language model	情感分析任务
WKLM	ICLR2020	Transformer Enc	对比学习	将文本中的实体使用同类实体替代	Masked Language Model	事实补全+QA+实体分类
Know BERT	ENNLP2019	Transformer Enc+ middleware	知识库固定+ 辅助结构	在BERT中引入中间件将知识库的嵌入表示融合进词嵌入	Masked Language Model+ Next Sentence Predict	关系抽取+实体识别
K-Adapter	EMNLP2019	Transformer Enc+ Independent Adapter	知识库固定+ 辅助结构	为每个独立的知识库训练一个adapter	relation classification+ Linguistic prediction (predicting the father index of each token)	实体分类+关系分类
K-BERT	AAAI2020	Transformer Enc+ Soft-position embedding+ Mask Attention	知识库固定+ 知识库嵌入文本	微调阶段将知识库实体和关系编码到文本中	Masked Language Model+ Next Sentence Predict	NLU+NER
Co-LAKE	COLING2020	Transformer Enc+ Soft-position embedding+ Mask Attention	知识库固定+ 知识图谱子图	训练阶段将知识库实体和关系编码到文本中	Masked Language Model (Masking word nodes+ Masking entity nodes+ Masking relation nodes)	实体识别+知识推理+知识图谱补全
GLM（Graph-guided Masked Language Model）	arXiv2020	Transformer Enc+ Distractor-Suppressed Ranking Block	预训练任务（实体MLM+ 实体ranking）	引入实体级别的负采样学习	Entity-Level Masking+ Distractor-Suppressed Ranking	QA+知识库补全
KEPLER	arXiv2019	Transformer Enc+ Knowledge Embedding	联合训练	同时训练KE和LM	KE loss +MLM loss	关系分类+实体分类+知识推理
JACKET	arXiv2020	BERT+GCN	联合训练	同时训练KE和LM	Entity category prediction+Relation type prediction+ Masked token prediction+Masked entity prediction	关系分类+实体分类+QA（基于知识库）