

K-Adapter(arxiv 2020.10)

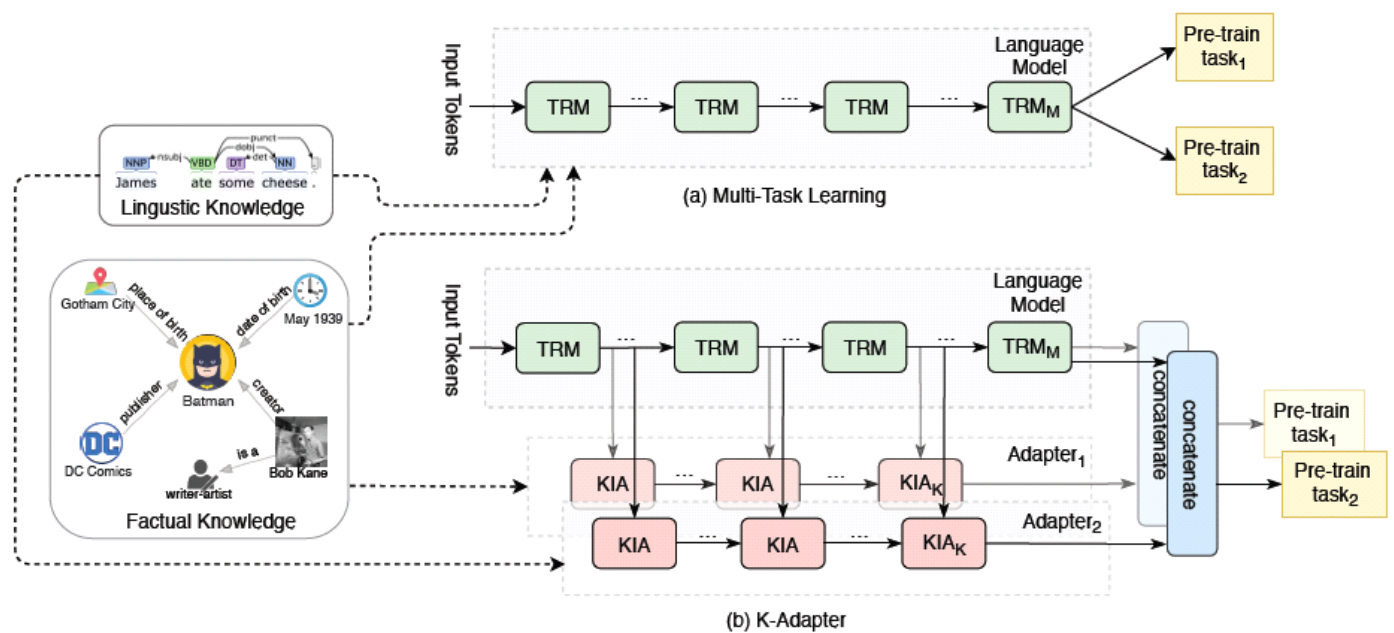
2020年11月20日 19:44

文章的主要动机:

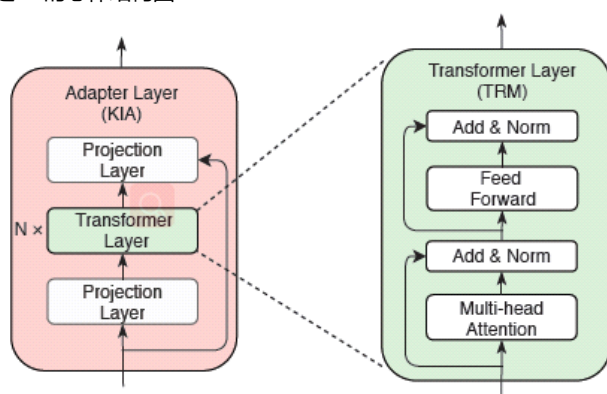
- (1) 将特定领域的知识在不改变模型参数的情况下, 融入一些稳定的预训练模型 (BERT,roBERTa) .
- (2) 为不同领域的知识单独创建一个Adapter, 作者认为如果将所有的知识不分类输入, 会导致无法评估特定的知识对模型的有效性 (个人不太认同这个观点, 混合的知识输入对于知识的迁移以及交叉领域可能有用) 。

Model	Knowledge Source	Objective	BERT fixed in training?	Continual knowledge infusion?
ERNIE (Zhang et al., 2019)	Wikipedia, WikiData	entity linking	N	N
LIBERT (Lauscher et al., 2019)	WordNet	synonym word prediction, hyponym-hypernym prediction	from scratch	N
SenseBERT (Levine et al., 2019)	WordNet	word-supersense prediction	from scratch	N
KnowBERT (Peters et al., 2019)	Wordnet, Wikipedia, CrossWikis	entity linking , hypernym linking	N	N
WKLM (Xiong et al., 2020)	WikiPedia, WikiData	replaced entity detection	N	N
BERT-MK (He et al., 2019)	Unified Medical Language System	discriminate between real and fake facts	N	N
K-Adapter (this work)	Wikipedia, Wikidata, dependency parser	predication prediction, dependency relation prediction	Y	Y

模型总体结构 (传统的多任务学习和k-adapter的对比)



下面是KIA的总体结构图



实验（重点）

(1) roBERTa(large)作为实验的基础模型，设置KIA的Transformer layer层N=2，下层Projection layer的隐变量H=768，Multi-head=12，上层projection layer H=1024,Transformer的H也是768.

(2) FACTUAL ADAPTER

事实知识的apapter训练方式：从T-REx中抽取了一个子集T-REx-rc，抛弃包含人名的50个实体对，收集了430个关系和5.5M个句子。然后输入句子，使用实体关系分类的方式训练adapter

(3) LINGUISTIC ADAPTER

使用Stanford Parser在Book Corpus上构建1M个依存关系句子，训练任务是预测父节点的token