***Time-aware language model based on continual/transfer learning***

*ZHIHAN ZHANG*

**Recent idea:**

* 基于time-aware LMs、CKL等方法，比较T5-Kadapters、T5-LoRA、T5 with prefix tuning “year: 2014”模型的time-aware ability
* 三者分别对应Parameter-efficient transfer learning中的三种架构Adapter、LoRA和Prefix tuning，前两者也是CKL中表现最佳的模型
* 潜在的创新点：组合上述框架，比如T5- LoRA with prefix tuning “year: 2014” ，可能得到更强的time-aware ability
* 目前正在复现CKL和T5 with prefix tuning “year: 2014”两篇论文，希望可以在实践中将idea更加细化

[[2110.04366] Towards a Unified View of Parameter-Efficient Transfer Learning (arxiv.org)](https://arxiv.org/abs/2110.04366)

**Parameter-efficient transfer learning**主要包括Adapter、Prefix tuning和LoRA

主要结论：good designs的特点是

* Inserting Form: Parallel > Sequential
* Modified Representation: FFN > attn (generally), but multi-head attn is superior with very small parameter budget (0.1% of original parameters)
* Composition: Scale addition is a good trade-off between performance and simplicity

**Google研究的Time-Aware LMs**

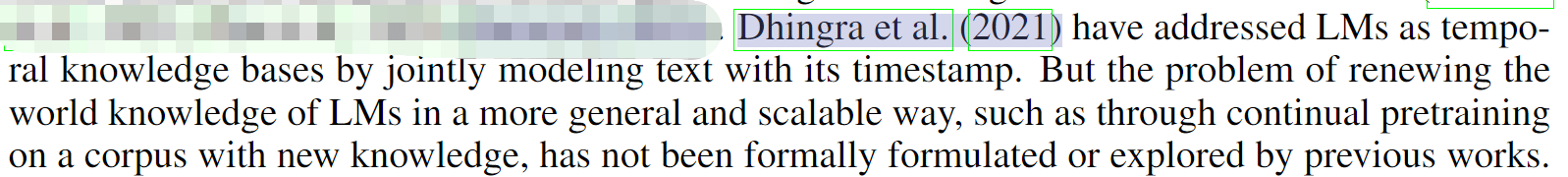
[Time-Aware Language Models as Temporal Knowledge Bases - ACL Anthology](https://aclanthology.org/2022.tacl-1.15/)

* T5 with prefix tuning “year: 2014”在文中被简称为Temporal，另外准备Yearly和Uniform作为对比模型，发现Temporal在效果上time-aware的任务上都优于Yearly和Uniform。
* 总之，T5 with prefix tuning “year: 2014”是表现很不错的Time-aware LMs

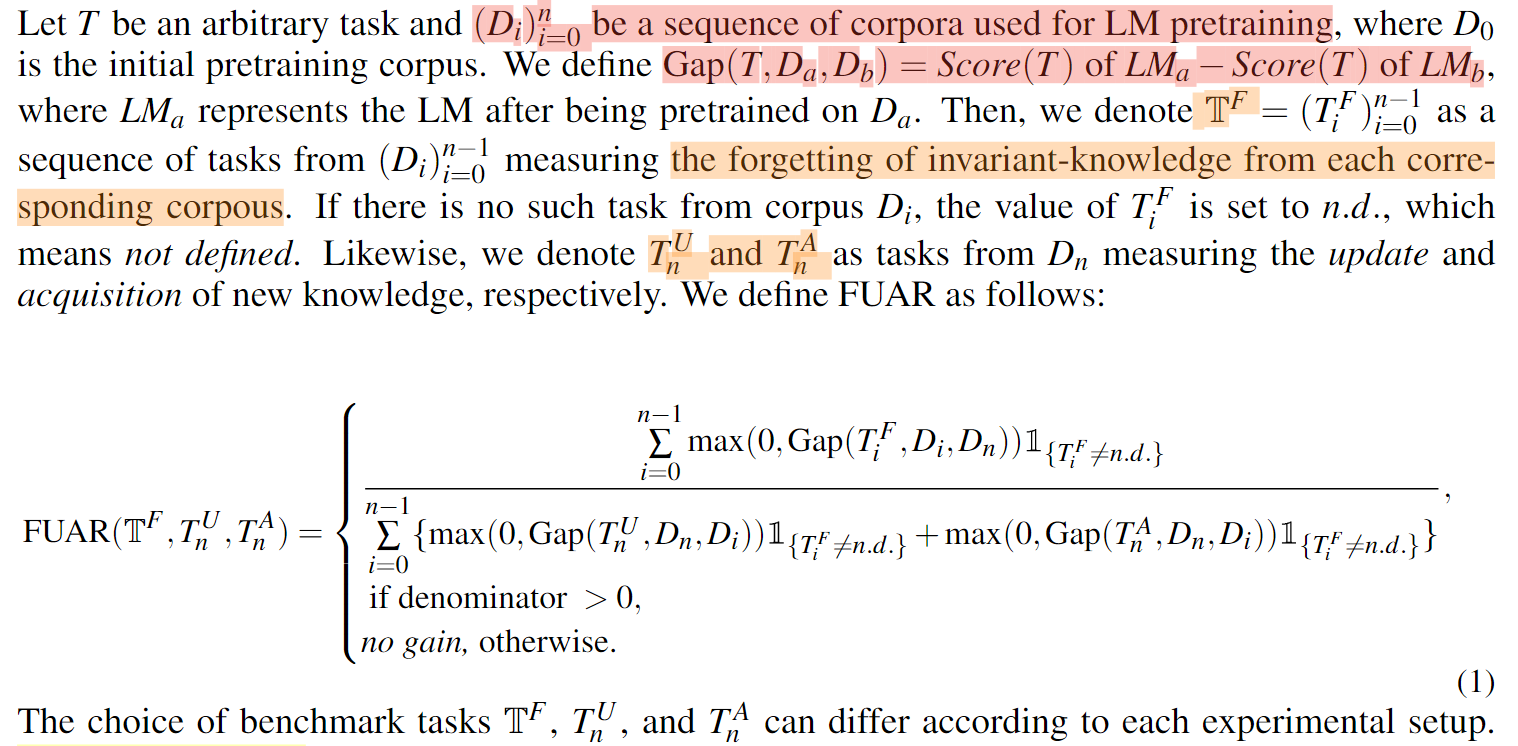
**KAIST AI研究的Continual Knowledge Learning**

[[2110.03215] Towards Continual Knowledge Learning of Language Models (arxiv.org)](https://arxiv.org/abs/2110.03215)

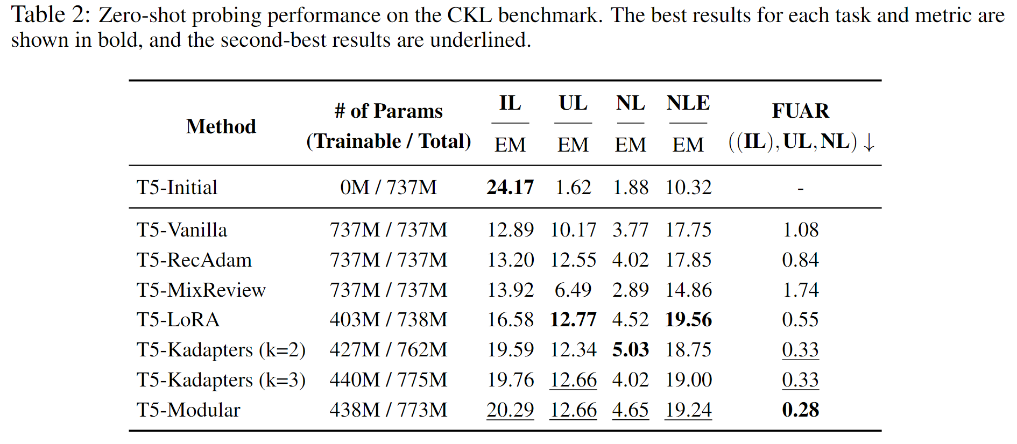
* KAIST AI此前制作了TemporalWiki，将Wikipedia数据分为Changed和Unchanged，直观理解应该是将LM持续在Changed数据上训练。Evaluation阶段，baseline models包括Initial、Full、Diff、Recdam、Mix-review、LoRA、K-Adapter，这种LM在NON-TWIKI-DIFFSETS的表现明显弱于continue/transfer learning，但在TWIKI-DIFFSETS表现最好。先不谈比较结果如何，KAIST AI正在将transfer learning的各种框架用于创造最具time-aware ability的LM
* TemporalWiki同时涉及了GPT2和T5，但在正文更加侧重GPT2；虽然CKL也同时涉及了GPT2和T5，但在正文更加侧重T5。
  + *为什么T5变得更popular呢？有时候看论文感到迷惑的点是，作者并不解释为什么选择这种模型，比如TwitterLM选择RoBERTa、TemporalWiki选择GPT2和T5。有没有卡，在不在乎判别任务*
* CKL提到了Google的工作，认为T5 with prefix tuning “year: 2014”的缺点在于不够general或scalable，表明自己的观点后并没有继续深入研究prefix tuning



* Continual knowledge learning的模型包括T5-RecAdam、T5-MixReview、T5-LoRA、T5-Kadapters、T5-Modular，其中RecAdam是regularization method，Mix-Review是rehearsal method，LoRA、K-Adapter、Modular是parameter-expansion method
  + *Modular没有对应的reference，不清楚作者是从哪里得到的newly proposed method，或许是作者自己提出来的？*
* 提出了Combined Metric for CKL，**FUAR** (**F**orgottrn / (**U**pdated+**A**cquired) **R**atio)，比较the efficiency of each CKL method using the trade-off between forgotten time-invariant knowledge and updated or newly acquired knowledge
  + FUAR=1.0意味着，模型每gain或update一个知识，它就会遗忘一个旧知识
  + FUAR <1.0意味着，模型获得新知识或更新旧知识的速度高于遗忘过去的知识



* 实验结果表明：
  + 在保留旧知识上（IL），T5-Initial表现最优，因为它从未被continued pretrained；其次是T5-Modular
  + 在更新旧知识上（UL），T5-LoRA、T5-Kadapters和T5-Modular表现很好，它们都是parameter-expansion method
  + 在学习新知识上（NL），仍然是parameter-expansion method表现很好
  + FUAR分数表明，T5 with parameter-expansion method的分值最低，意味着它们能在获得新知识和更新旧知识的同时，最少地忘记旧知识
  + **总之，T5 with parameter-expansion method的模型表现最好，最适合continual learning**



**Some Conclusion**

* Google研究的Time-Aware LMs使用了TL中的prefix tuning方法，KAIST AI研究的Continual Knowledge Learning使用了TL中的Adapter、LoRA等parameter-expansion方法。
* T5 with prefix tuning在TEMPLAMA上表现优于分年份训练出的九个Yearly模型，T5 with parameter-expansion method在FUAR上表现优于T5 with regularization method和T5 with rehearsal method
* 相比之下，KAIST AI认为T5 with parameter-expansion method的优势在于，它比T5 with prefix tuning更加general或scalable。但是这一说法并不能说明T5 with parameter-expansion method在time-aware ability上真的做的更好，它不一定在TEMPLAMA上的表现优于T5 with prefix tuning
* 我希望可以用实验比较T5 with prefix tuning和T5 with parameter-expansion method的time-aware ability，并且看是否可以结合两种方法提出一种更time-aware的模型