PALT: Parameter-Lite Transfer of Language Models for Knowledge Graph Completion

Findings of EMNLP 2022

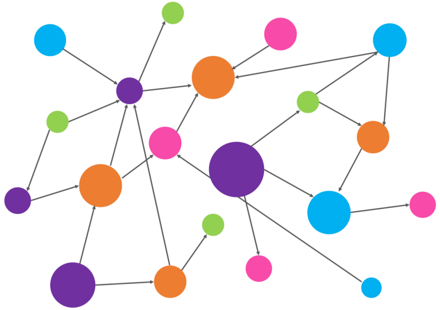
Jianhao Shen, Chenguang Wang, Ye Yuan, Jiawei Han, Heng Ji, Koushik Sen, Ming Zhang, Dawn Song

Knowledge Graph

Knowledge Graph

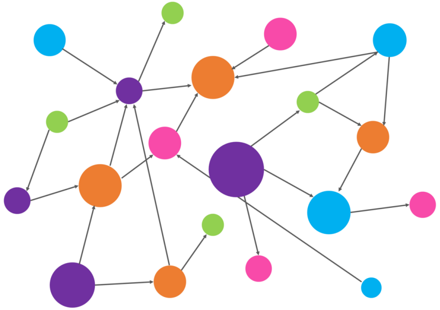
Form

(head entity, relation, tail entity)
(Steve Jobs, founder of, Apple Inc.)



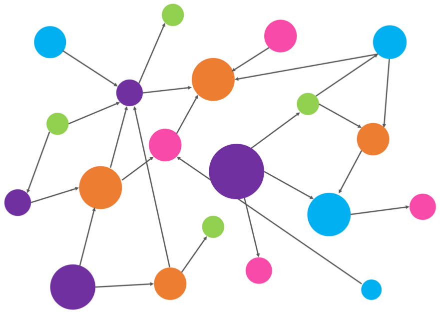
Knowledge Graph

Form (head entity, relation, tail entity)
(Steve Jobs, founder of, Apple Inc.)



Instances Freebase WORDNET Unified Medical Language System

Knowledge Graph



Form

(head entity, relation, tail entity)
(Steve Jobs, founder of, Apple Inc.)

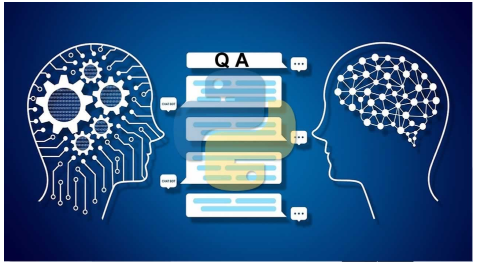
Instances

Freebase

WORDNET

Unified Medical Language System

Application



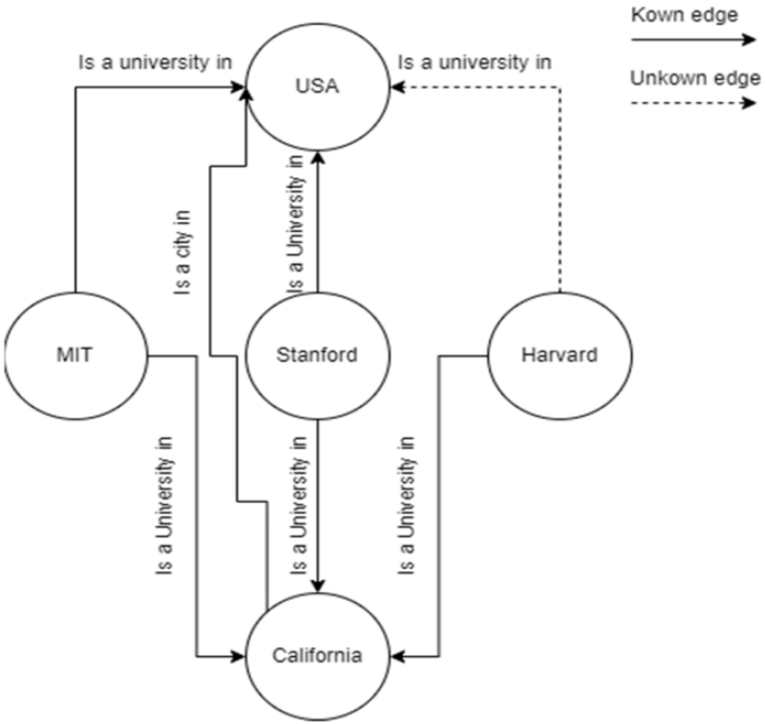
Knowledge graph has been applied to many applications

Knowledge Graph Completion

- Real-world knowledge graphs are usually incomplete

Knowledge Graph Completion

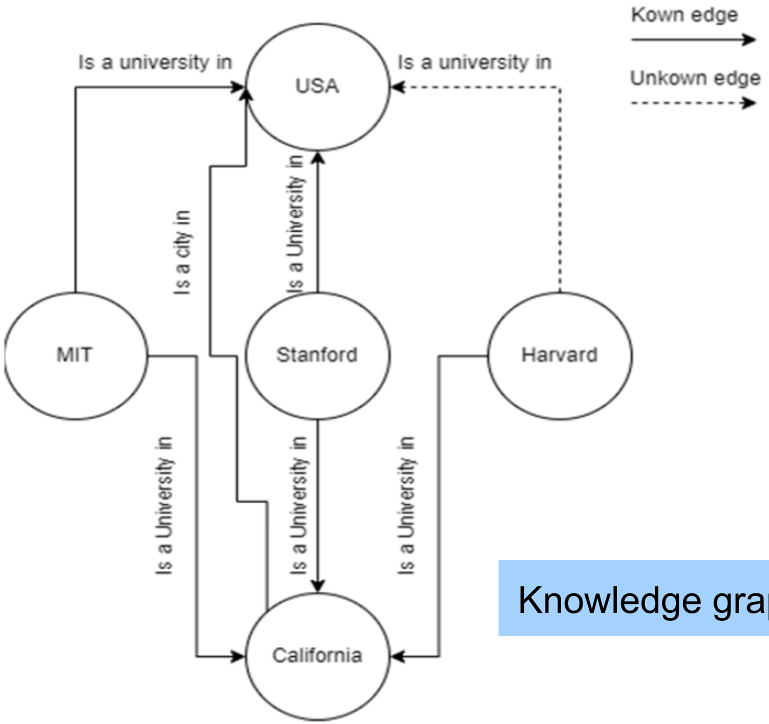
- Real-world knowledge graphs are usually incomplete



(Harvard, ?, USA)

Knowledge Graph Completion

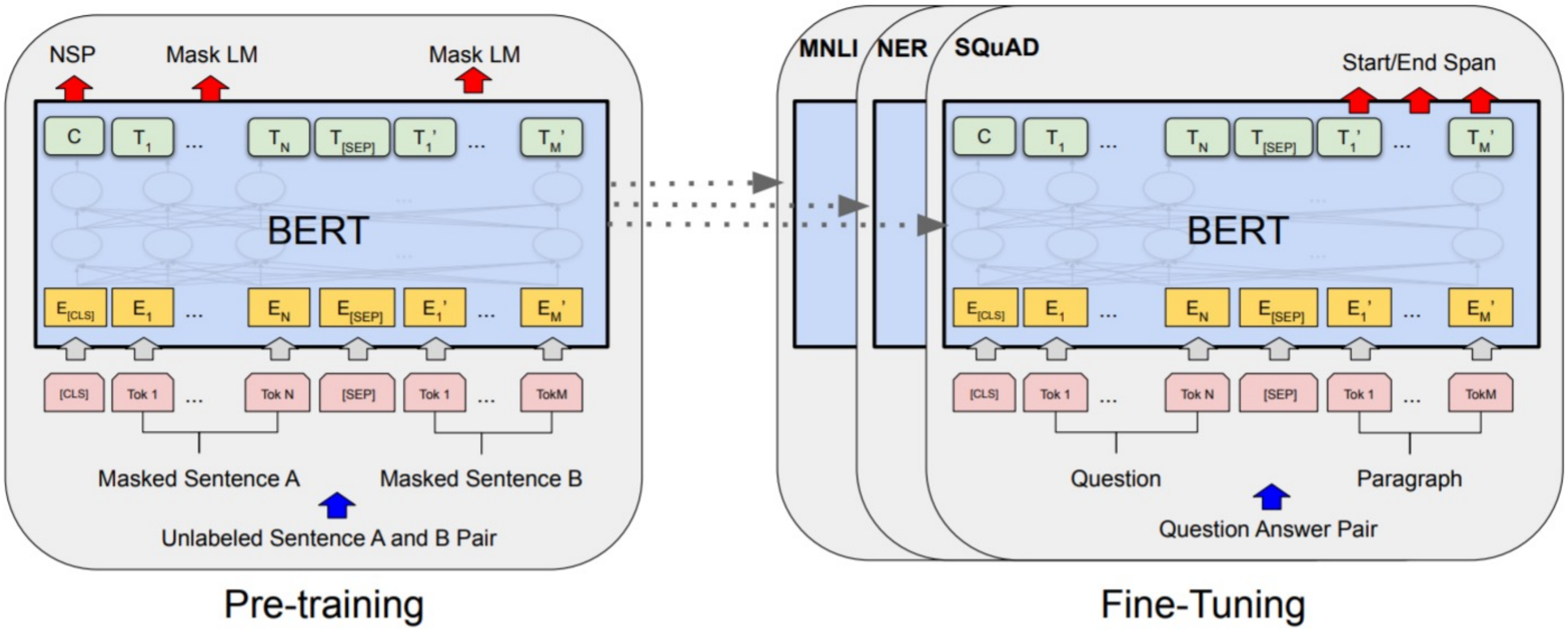
- Real-world knowledge graphs are usually incomplete



(Harvard, ?, USA)

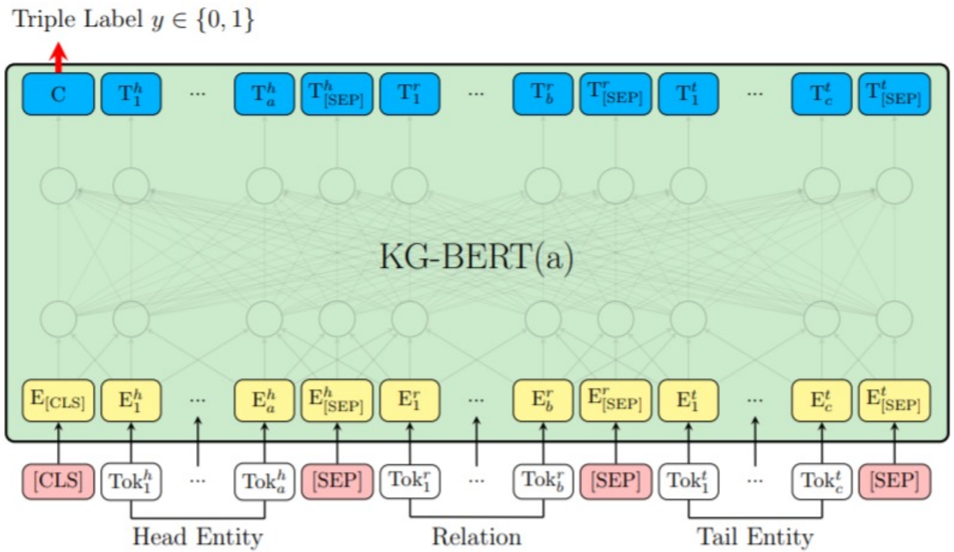
Knowledge graph completion is important for real-world applications

Pretrained Language Models



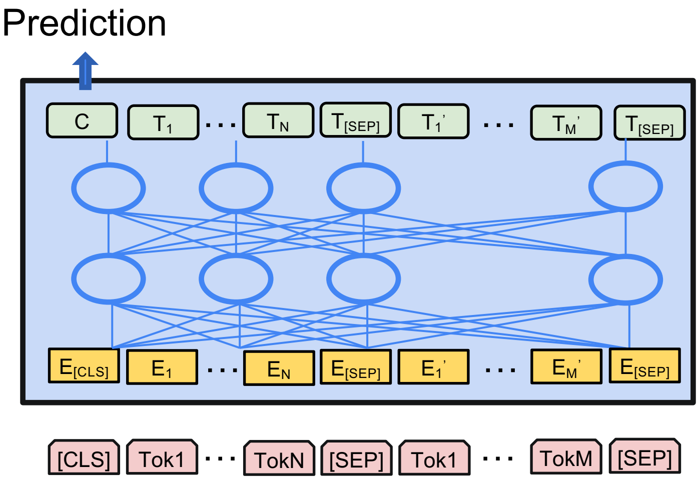
Pretrained language models have enabled downstream transfer

Finetuning to Knowledge Graph Completion

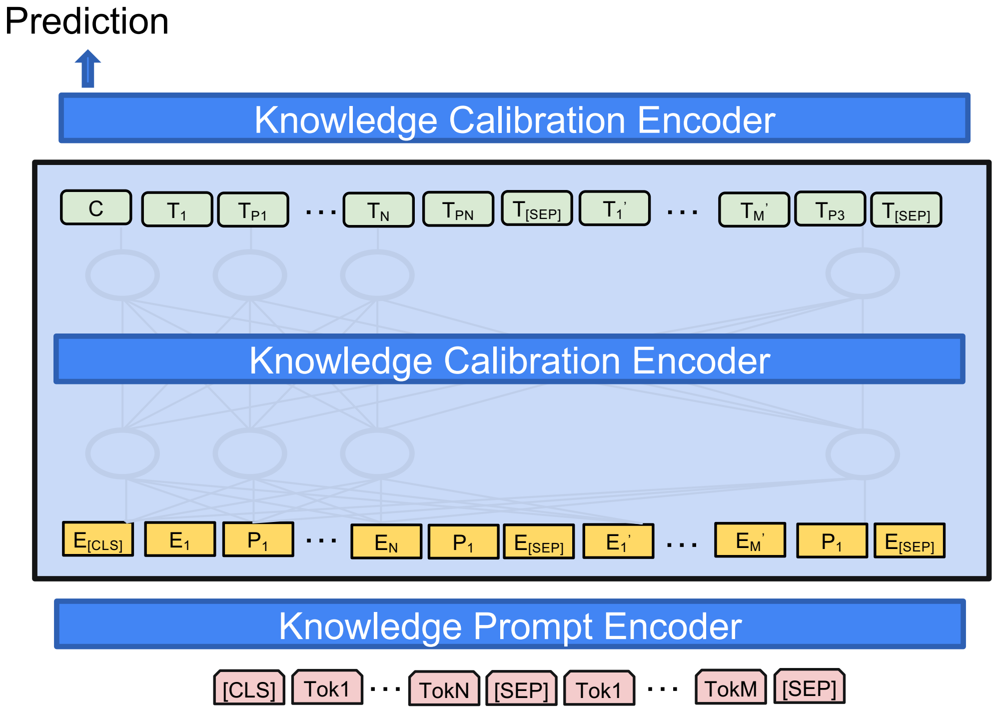


Finetuning requires high computational and storage resources

PArameter-**L**ite **T**ransfer (PALT)



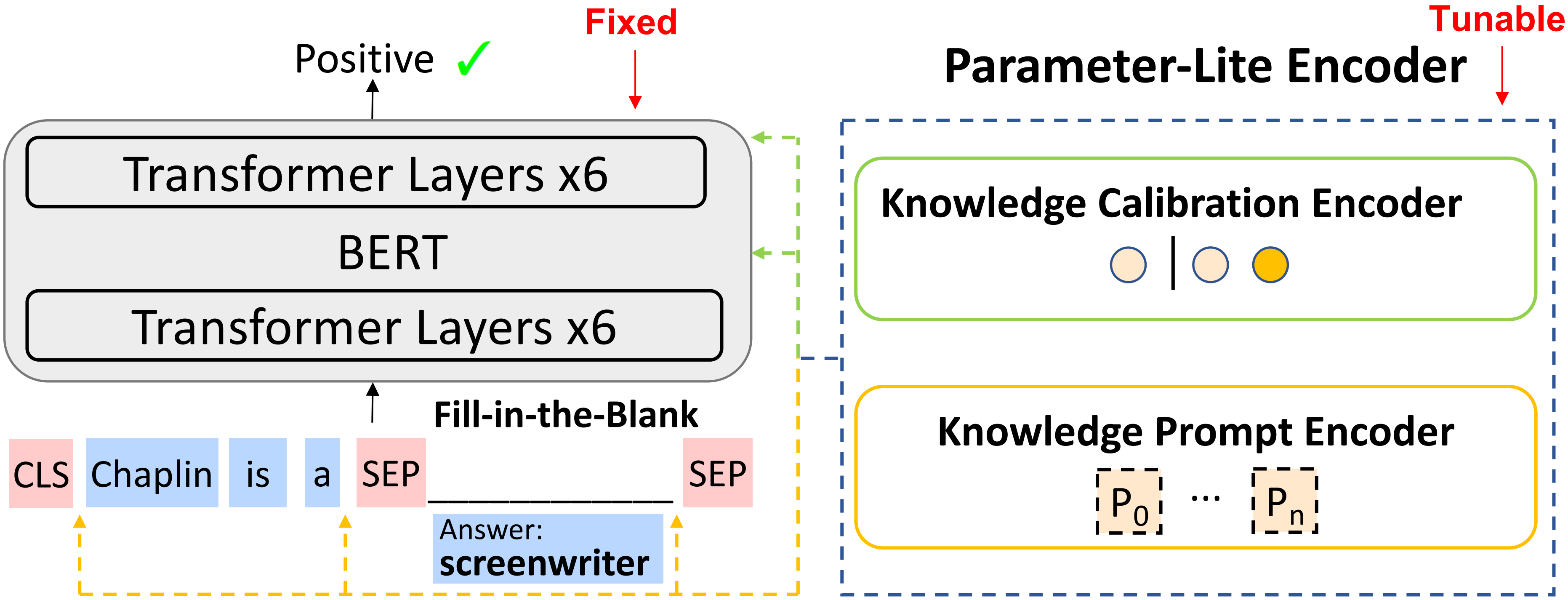
Finetuning



PALT

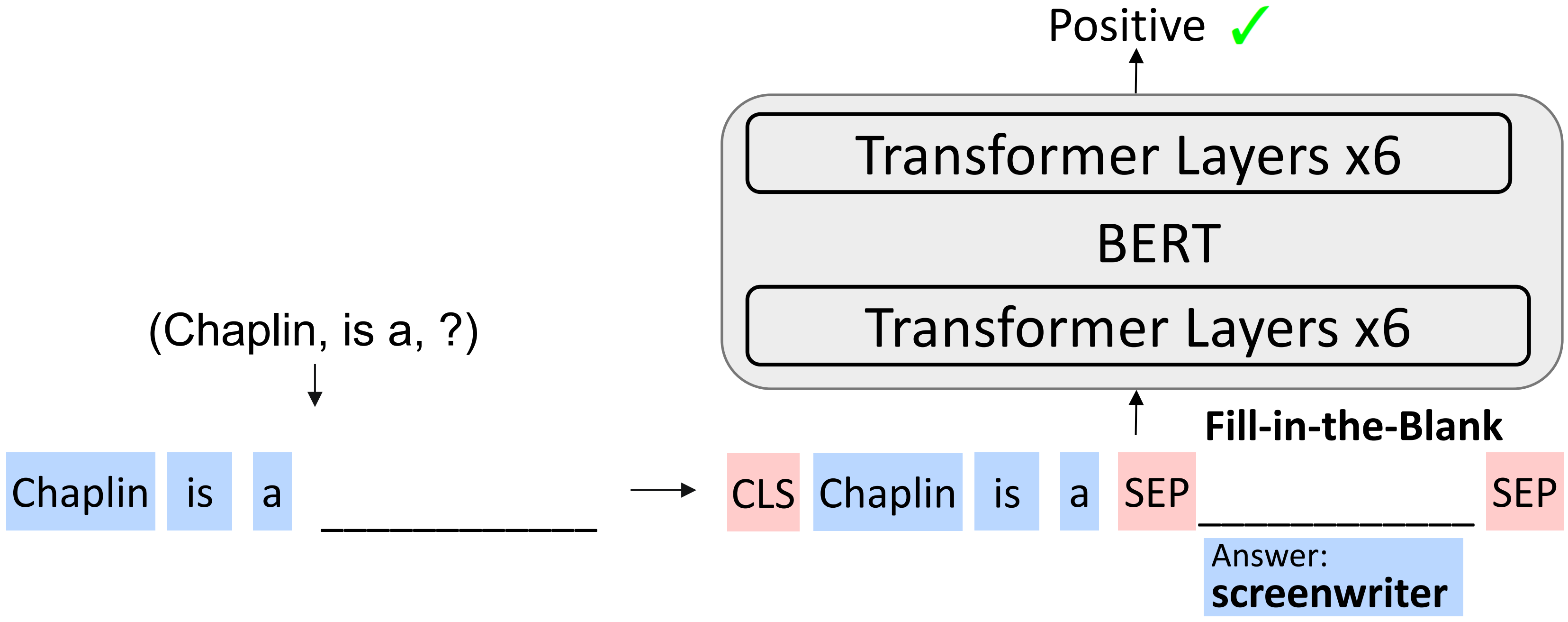
Intuition: not all parameters need to be tuned for transfer

PArameter-Lite Transfer (PALT)



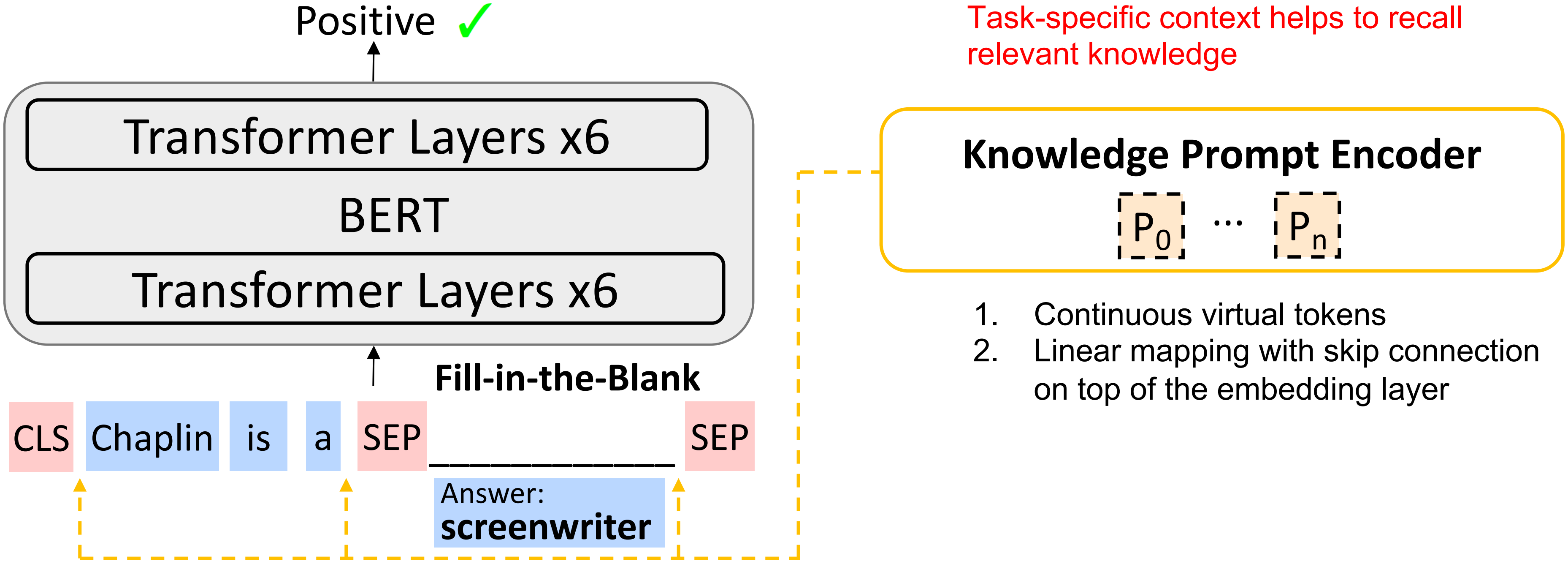
Prepare a proper context to recall knowledge and align the output distribution with the task of interest

Task Formulation



Cloze-style statements trigger language models to produce general knowledge

Knowledge Prompt Encoder

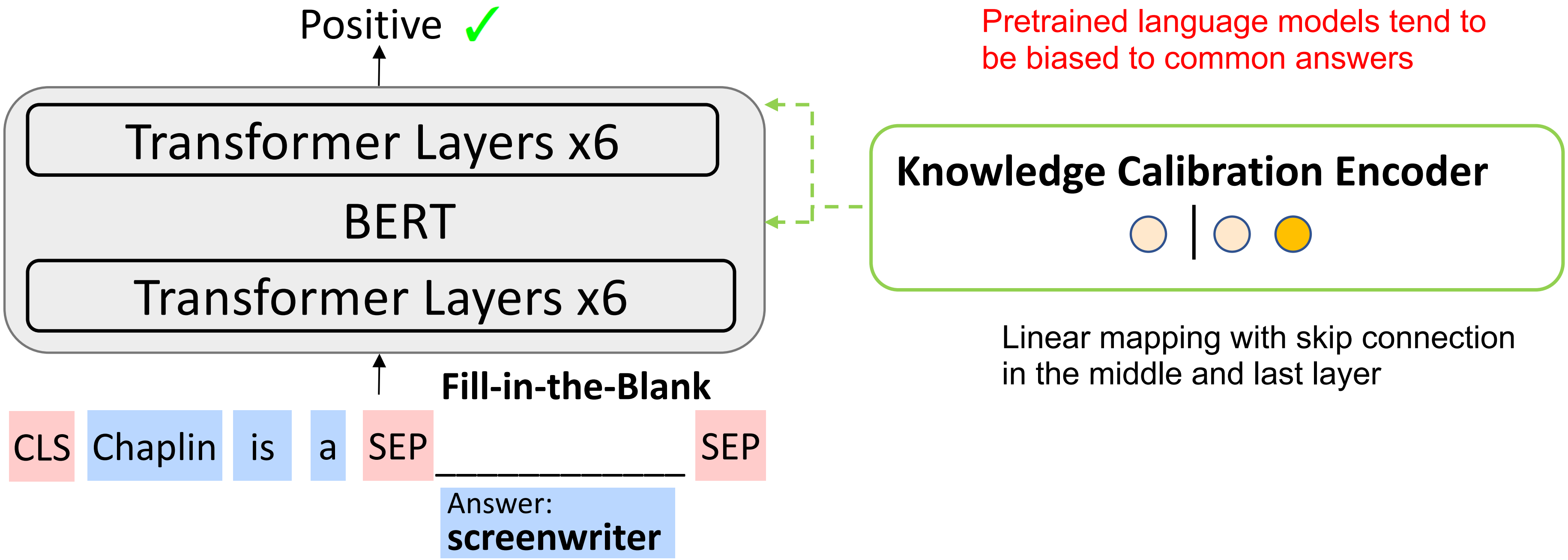


Task-specific context helps to recall relevant knowledge

1. Continuous virtual tokens
2. Linear mapping with skip connection on top of the embedding layer

Knowledge prompt encoder provides task-specific context

Knowledge Calibration Encoder



Pretrained language models tend to be biased to common answers

Linear mapping with skip connection in the middle and last layer

Knowledge calibration encoder aligns the pretraining distribution with downstream distribution

Training

- Only tune parameter-lite encoders and keep PLM parameters fixed
- Training objective:
  - NSP prediction about whether two sentences are correctly connected

NSP classifier

$$L_h = -\log \Pr(1|h, r, t) - \sum_{i}^{n_{\text{ns}}} \mathbb{E}_{\tilde{h}_i \sim E \setminus \{h\}} \log \Pr(0|\tilde{h}_i, r, t)$$

Positive samples

Negative samples

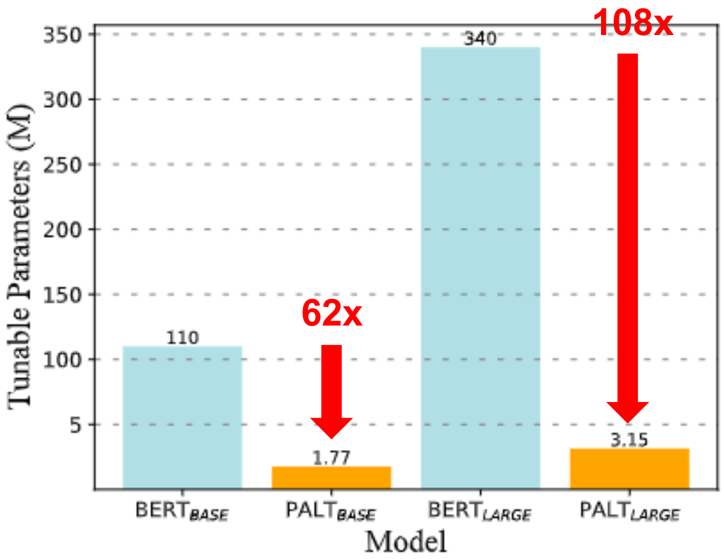
Experiment

Method	WN11	FB13	Avg
Task-specific models			
NTN (Socher et al., 2013)	86.2	90.0	88.1
TransE (Bordes et al., 2013)	75.9	81.5	78.7
TransH (Wang et al., 2014)	78.8	83.3	81.1
TransR (Lin et al., 2015)	85.9	82.5	84.2
TransD (Ji et al., 2015)	86.4	89.1	87.8
TEKE (Wang and Li, 2016)	86.1	84.2	85.2
TransG (Xiao et al., 2016)	87.4	87.3	87.4
TranSparse-S (Ji et al., 2016)	86.4	88.2	87.3
DistMult (Yang et al., 2015)	87.1	86.2	86.7
DistMult-HRS (Zhang et al., 2020a)	88.9	89.0	89.0
AATE (An et al., 2018)	88.0	87.2	87.6
ConvKB (Nguyen et al., 2018)	87.6	88.8	88.2
DOLORES (Wang et al., 2020b)	87.5	89.3	88.4
General models			
KG-BERT (Yao et al., 2019)	93.5	90.4	91.9
$\text{PALT}_{\text{BASE}}$ (ours)	93.3	91.3	92.3
$\text{PALT}_{\text{LARGE}}$ (ours)	93.8	91.7	92.8

Triplet Classification

Our approach achieves state-of-the-art performance and is better than fully finetuning methods

Number of Parameters



PALT achieves better performance than finetuning with only 1% tunable parameters

Ablation Study

Method	WN11
$\text{PALT}_{\text{BASE}}$	93.3
w/o Prompt	91.7
w/o $\text{Calibration}_{\text{middle}}$	92.2
w/o $\text{Calibration}_{\text{last}}$	93.0
w/o $\text{Calibration}_{\text{both}}$	89.3
w/o Encoder	73.7
Finetuning	93.2

All components have a positive effect

Conclusion

- We propose parameter-lite transfer (PALT) for knowledge graph completion
    - Task formulation is important to use knowledge in language models
    - Novel prompt encoders and calibration encoders which can be used for other models
- PALT achieves state-of-the-art performance
- It should be useful for broad knowledge-intensive NLP applications
- Knowledge graph completion is a knowledge benchmark for language models

Thank you for your time!

Code:https://github.com/yuanyehome/PALT